

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
22 March 2001 (22.03.2001)

PCT

(10) International Publication Number  
**WO 01/20004 A2**(51) International Patent Classification<sup>7</sup>: **C12N 15/55**,  
15/54, 15/11, 9/12, 9/16, C07K 16/40, C12Q 1/68, A01K  
67/027, A61K 31/713, G01N 33/50, 33/53Mountain View, CA 94040 (US). **BAUGHN, Mariah, R.**  
[US/US]; 14244 Santiago Road, San Leandro, CA 94577  
(US). **AZIMZAI, Yalda** [US/US]; 2045 Rock Springs  
Drive, Hayward, CA 94545 (US). **LU, Dyung, Aina, M.**  
[US/US]; 55 Park Belmont Place, San Jose, CA 95136  
(US).

(21) International Application Number: PCT/US00/25515

(22) International Filing Date:

14 September 2000 (14.09.2000)

(74) Agents: **HAMLET-COX, Diana et al.**; Incyte Genomics,  
Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/154,141 15 September 1999 (15.09.1999) US

(63) Related by continuation (CON) or continuation-in-part  
(CIP) to earlier application:US 60/154,141 (CIP)  
Filed on 15 September 1999 (15.09.1999)(71) Applicant (for all designated States except US): **INCYTE  
GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo  
Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **YUE, Henry**  
[US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US).  
**TANG, Y., Tom** [CN/US]; 4230 Ranwick Court, San  
Jose, CA 95118 (US). **BANDMAN, Olga** [US/US]; 366  
Anna Avenue, Mountain View, CA 94043 (US). **HILL-  
MAN, Jennifer, L.** [US/US]; 230 Monroe Drive, #17,(81) Designated States (national): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ,  
DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,  
HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,  
LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,  
NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,  
TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.(84) Designated States (regional): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian  
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European  
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,  
IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG,  
CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).**Published:**— Without international search report and to be republished  
upon receipt of that report.For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.

(54) Title: PROTEIN PHOSPHATASE AND KINASE PROTEINS

(57) Abstract: The invention provides human protein phosphatase and kinase proteins (PPHKP) and polynucleotides which identify and encode PPHKP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of PPHKP.

WO 01/20004 A2

## PROTEIN PHOSPHATASE AND KINASE PROTEINS

### TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of protein phosphatase and kinase proteins and to the use of these sequences in the diagnosis, treatment, and prevention of gastrointestinal disorders, immune system disorders, neurological disorders, and cell proliferative disorders, including cancer.

### BACKGROUND OF THE INVENTION

Kinases and phosphatases are critical components of intracellular signal transduction mechanisms. Kinases catalyze the transfer of high energy phosphate groups from adenosine triphosphate (ATP) to hydroxyamino acids of various target proteins. Phosphatases, in contrast, remove phosphate groups from proteins. Reversible protein phosphorylation is the main strategy for regulating protein activity in eukaryotic cells. In general, proteins are activated by phosphorylation in response to extracellular signals such as hormones, neurotransmitters, and growth and differentiation factors. Protein dephosphorylation occurs when down-regulation of a signaling pathway is required. The combined activities of kinases and phosphatases regulate key cellular processes such as proliferation, differentiation, and cell cycle progression.

#### Protein Kinases

Kinases comprise the largest known enzyme superfamily and vary widely in their target proteins. Kinases may be categorized as protein tyrosine kinases (PTKs), which phosphorylate tyrosine residues, and protein serine/threonine kinases (STKs), which phosphorylate serine and/or threonine residues. Some kinases have dual specificity for both serine/threonine and tyrosine residues. Almost all kinases contain a conserved 250-300 amino acid catalytic domain. This domain can be further divided into 11 subdomains. N-terminal subdomains I-IV fold into a two-lobed structure which binds and orients the ATP donor molecule, and subdomain V spans the two lobes. C-terminal subdomains VI-XI bind the protein substrate and transfer the gamma phosphate from ATP to the hydroxyl group of a serine, threonine, or tyrosine residue. Each of the 11 subdomains contains specific catalytic residues or amino acid motifs characteristic of that subdomain. For example, subdomain I contains an 8-amino acid glycine-rich ATP binding consensus motif, subdomain II contains a critical lysine residue required for maximal catalytic activity, and subdomains VI through IX comprise the highly conserved catalytic core. STKs and PTKs also contain distinct sequence motifs in subdomains VI and VIII which may confer hydroxyamino acid specificity. Some STKs and PTKs possess structural characteristics of both families. In addition, kinases may also be classified by additional amino acid sequences, generally between 5 and 100 residues, which either flank or

occur within the kinase domain. These additional amino acid sequences regulate kinase activity and determine substrate specificity. (Reviewed in Hardie, G. and Hanks, S. (1995) The Protein Kinase Facts Book, Vol I:7-20 Academic Press, San Diego, CA.)

The second messenger dependent protein kinases primarily mediate the effects of second messengers such as cyclic AMP (cAMP), cyclic GMP, inositol triphosphate, phosphatidylinositol, 3,4,5-triphosphate, cyclic ADPribose, arachidonic acid, diacylglycerol and calcium-calmodulin. The cyclic-AMP dependent protein kinases (PKA) are important members of the STK family. Cyclic-AMP is an intracellular mediator of hormone action in all animal cells that have been studied. Hormone-induced cellular responses include thyroid hormone secretion, cortisol secretion, progesterone secretion, glycogen breakdown, bone resorption, and regulation of heart rate and force of heart muscle contraction. PKA is found in all animal cells and is thought to account for the effects of cyclic-AMP in most of these cells. Altered PKA expression is implicated in a variety of disorders and diseases including cancer, thyroid disorders, diabetes, atherosclerosis, and cardiovascular disease (Isselbacher, K.J. et al. (1994) Harrison's Principles of Internal Medicine, McGraw-Hill, New York, NY, pp. 416-431, 1887).

Calcium-calmodulin (CaM) dependent protein kinases are also members of the STK family. Calmodulin is a calcium receptor that mediates many calcium regulated processes by binding to target proteins. The principle target proteins in these processes are CaM-dependent protein kinases (CaM kinases). CaM kinases are involved in regulation of smooth muscle contraction, glycogen breakdown (phosphorylase kinase), and neurotransmission (CaM kinase I and CaM kinase II). CaM kinase I phosphorylates a variety of substrates including the neurotransmitter-related proteins synapsin I and II, the gene transcription regulator, CREB, and the cystic fibrosis conductance regulator protein, CFTR (Haribabu, B. et al. (1995) *EMBO Journal* 14:3679-3686). CaM kinase II also phosphorylates synapsin at different sites and controls the synthesis of catecholamines in the brain through phosphorylation and activation of tyrosine hydroxylase. Many CaM kinases are activated not only by CaM, but also by phosphorylation. For example, CaM kinase may autophosphorylate itself or may be phosphorylated by another kinase as part of a kinase cascade. The mRNA encoding a calmodulin-binding protein kinase-like protein was found to be enriched in mammalian forebrain. This protein is associated with vesicles in both axons and dendrites and accumulates largely postnatally. The amino acid sequence of this protein is similar to CaM-dependent STKs, and the protein binds calmodulin in the presence of calcium (Godbout, M. et al. (1994) *J. Neurosci.* 14:1-13).

The mitogen-activated protein kinases (MAP) are another STK family that regulates intracellular signaling pathways. The MAP kinases mediate signal transduction from the cell surface to the nucleus via phosphorylation cascades. Several subgroups have been identified, and each manifests different substrate specificities and responds to distinct extracellular stimuli (Egan, S.E.

and Weinberg, R.A. (1993) *Nature* 365:781-783). MAP kinase signaling pathways are present in mammalian cells as well as in yeast. The extracellular stimuli which activate MAP kinase pathways include epidermal growth factor (EGF), ultraviolet light, hyperosmolar medium, heat shock, endotoxic lipopolysaccharide (LPS), and pro-inflammatory cytokines such as tumor necrosis factor (TNF) and interleukin-1 (IL-1). Altered MAP kinase expression is implicated in a variety of disease conditions including cancer, inflammation, immune disorders, and disorders affecting growth and development.

PTKs may be divided into transmembrane, receptor PTKs and nontransmembrane, non-receptor PTKs. Transmembrane PTKs are receptors for most growth factors. Binding of growth factor to the receptor activates the transfer of a phosphate group from ATP to selected intracellular tyrosine side chains of the receptor and other intracellular signal transduction proteins. Growth factors (GF) which bind to receptor PTKs include epidermal GF, platelet-derived GF, fibroblast GF, hepatocyte GF, insulin and insulin-like GFs, nerve GF, vascular endothelial GF, and macrophage colony stimulating factor. Non-receptor PTKs lack transmembrane regions and, instead, form complexes with the intracellular regions of cell surface receptors. Receptors that function through non-receptor PTKs include those for cytokines and hormones (growth hormone and prolactin), and antigen-specific receptors on T and B lymphocytes.

Many PTKs were first identified as the products of mutant oncogenes in cancer cells in which their activation was no longer subject to normal cellular controls. In fact, about one third of the known oncogenes encode PTKs, and it is well known that cellular transformation (oncogenesis) is often accompanied by increased tyrosine phosphorylation activity (Charbonneau H. and Tonks N.K. (1992) *Annu Rev Cell Biol* 8:463-493). Regulation of PTK activity may therefore be an important strategy in controlling some types of cancer.

#### Protein Phosphatases

Phosphatases are characterized as either tyrosine-specific or serine/threonine-specific based on their preferred phospho-amino acid substrate. However, some phosphatases exhibit dual specificity for both phospho-tyrosine and phospho-serine/threonine residues. Many serine/threonine-specific phosphatases (STPs) have been biochemically purified and extensively characterized. STPs are generally comprised of two or more subunits and have broad and overlapping protein substrate specificities. STPs are found in the cytosol, nucleus, and mitochondria and in association with cytoskeletal and membranous structures. Some STPs require divalent cations, such as  $\text{Ca}^{2+}$  or  $\text{Mn}^{2+}$ , for activity. STPs play important roles in glycogen metabolism, muscle contraction, protein synthesis, oocyte maturation, and hepatic metabolism. (Reviewed in Cohen, P. (1989) *Annu. Rev. Biochem.* 58:453-508.)

Protein phosphatase 2A (PP2A) is a well-characterized STP purified primarily from

mammalian tissues. A trimeric PP2A holoenzyme has been purified from rabbit skeletal muscle.

(Hendrix, P. et al. (1993) J. Biol. Chem. 268:15267-15276.) PP2A holoenzyme activity is stimulated by polycationic macromolecules. PP2A holoenzyme consists of a single catalytic subunit of 36 kilodaltons (kDa) and two regulatory subunits of 65 and 72 kDa. The regulatory subunits appear to  
5 determine the substrate-specificity, enzymatic activity, and subcellular localization of the holoenzyme. The 72 kDa subunit (PR72), in particular, increases the stimulatory effect of polycations and is phosphorylated by casein kinases I and II. The cDNA encoding human PR72 has been cloned, and PR72 gene transcripts have been detected exclusively in skeletal muscle and heart, suggesting that PR72 confers tissue-specificity to the PP2A holoenzyme. A 55 kDa regulatory  
10 subunit was also purified from the trimeric form of protein phosphatase 2A from rabbit skeletal muscle. This PR55 subunit was found to be encoded by two genes, alpha and beta. A high degree of conservation was found in a comparison of the amino acid sequences of human and rabbit PR55. The human PR55 beta isoform was detected at high levels in a neuroblastoma-derived cell line and at very low levels in other human cell lines. This suggests the isoform is neuron-specific (Mayer, R.E. et al.  
15 (1991) Biochemistry 30:3589-3597).

In contrast to STPs, tyrosine-specific phosphatases (YPs) are generally monomeric proteins which function primarily in the transduction of signals across the plasma membrane. YPs are categorized as either transmembrane receptor-like proteins or soluble nontransmembrane proteins. YPs share a conserved catalytic domain of about 250 amino acids which contains the active site. The  
20 active site consensus sequence consists of 13 amino acids including a cysteine residue that is essential for phosphatase activity. YPs play important roles in lymphocyte activation and cell adhesion. In addition, the genes encoding at least eight YPs have been mapped to chromosomal regions that are translocated or rearranged in various neoplastic conditions, including lymphoma, leukemia, small cell lung carcinoma, adenocarcinoma, and neuroblastoma. (Reviewed in Charbonneau, H. and Tonks, N.  
25 K. (1992) Annu. Rev. Cell Biol. 8:463-493.)

The discovery of new protein phosphatase and kinase proteins and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of gastrointestinal disorders, immune system disorders, neurological disorders, and cell proliferative disorders, including cancer.

## SUMMARY OF THE INVENTION

The invention features purified polypeptides, protein phosphatase and kinase proteins, referred to collectively as "PPHKP" and individually as PPHKP-1," "PPHKP-2," "PPHKP-3," "PPHKP-4," "PPHKP-5," "PPHKP-6," "PPHKP-7," "PPHKP-8," "PPHKP-9," "PPHKP-10," and  
35 "PPHKP-11." In one aspect, the invention provides an isolated polypeptide comprising an amino acid

sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-11.

The invention further provides an isolated polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-11. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:12-22.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the

polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11.

The invention further provides an isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said

target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a

5 polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino

10 acid sequence selected from the group consisting of SEQ ID NO:1-11, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional PPHKP, comprising administering to a patient in need of such treatment the composition.

15 The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino

20 acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another

25 alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional PPHKP, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide comprising an amino acid sequence selected from the group consisting

30 of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. The

35 method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting

antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional PPHKP, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:12-22, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound;

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

### BRIEF DESCRIPTION OF THE TABLES

Table 1 shows polypeptide and nucleotide sequence identification numbers (SEQ ID NOs), clone identification numbers (clone IDs), cDNA libraries, and cDNA fragments used to assemble full-length sequences encoding PPHKP.

Table 2 shows features of each polypeptide sequence, including potential motifs, homologous sequences, and methods, algorithms, and searchable databases used for analysis of PPHKP.

Table 3 shows selected fragments of each nucleic acid sequence; the tissue-specific expression patterns of each nucleic acid sequence as determined by northern analysis; diseases, disorders, or conditions associated with these tissues; and the vector into which each cDNA was cloned.

Table 4 describes the tissues used to construct the cDNA libraries from which cDNA clones encoding PPHKP were isolated.

Table 5 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

## DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

## DEFINITIONS

"PPHKP" refers to the amino acid sequences of substantially purified PPHKP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of PPHKP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PPHKP either by directly interacting with PPHKP or by acting on components of the biological pathway in which PPHKP participates.

An "allelic variant" is an alternative form of the gene encoding PPHKP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding PPHKP include those sequences with deletions,

insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as PPHKP or a polypeptide with at least one functional characteristic of PPHKP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding PPHKP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding PPHKP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent PPHKP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of PPHKP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of PPHKP. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PPHKP either by directly interacting with PPHKP or by acting on components of the biological pathway in which PPHKP participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind PPHKP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin,

thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic PPHKP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding PPHKP or fragments of PPHKP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (PE Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
15	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
20	Glu	Asp, Gln, His
	Gly	Ala
	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
25	Lys	Arg, Gln, Glu
	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
30	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
	Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a

polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

5 A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

A "fragment" is a unique portion of PPHKP or the polynucleotide encoding PPHKP which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example,  
10 a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected  
15 from the first 250 or 500 amino acids (or first 25% or 50% of a polypeptide) as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:12-22 comprises a region of unique polynucleotide sequence that  
20 specifically identifies SEQ ID NO:12-22, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:12-22 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:12-22 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:12-22 and the region of SEQ ID NO:12-22 to which the fragment corresponds are routinely  
25 determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-11 is encoded by a fragment of SEQ ID NO:12-22. A fragment of SEQ ID NO:1-11 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-11. For example, a fragment of SEQ ID NO:1-11 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-11. The precise length of  
30 a fragment of SEQ ID NO:1-11 and the region of SEQ ID NO:1-11 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full-length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A  
35 "full-length" polynucleotide sequence encodes a "full-length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Reward for match: 1*

*Penalty for mismatch: -2*

*Open Gap: 5 and Extension Gap: 2 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 11*

*Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (Apr-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Open Gap: 11 and Extension Gap: 1 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 3*

*Filter: on*

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

“Human artificial chromosomes” (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term “humanized antibody” refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

“Hybridization” refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the “washing” step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present

invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C<sub>0</sub>t or R<sub>0</sub>t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of PPHKP which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of PPHKP which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of PPHKP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of PPHKP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide,

polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

5 "Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

10 "Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

15 "Post-translational modification" of a PPHKP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of PPHKP.

20 "Probe" refers to nucleic acid sequences encoding PPHKP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes.

"Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and 25 identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers 30 may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular 35 Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR

Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

5           Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the  
10   PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which  
15   sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments,  
20   thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to  
25   identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

          A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the  
30   artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to  
35   transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

5 A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent,  
10 chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose  
15 instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing nucleic acids encoding PPHKP, or fragments thereof, or PPHKP itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

20 The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide  
25 comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which  
30 they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers,  
35 microparticles and capillaries. The substrate can have a variety of surface forms, such as wells,

trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants, and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook, J. et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least 98% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the

reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

### THE INVENTION

The invention is based on the discovery of new human protein phosphatase and kinase proteins (PPHKP), the polynucleotides encoding PPHKP, and the use of these compositions for the diagnosis, treatment, or prevention of gastrointestinal disorders, immune system disorders, neurological disorders, and cell proliferative disorders, including cancer.

Table 1 lists the Incyte clones used to assemble full length nucleotide sequences encoding PPHKP. Columns 1 and 2 show the sequence identification numbers (SEQ ID NOs) of the polypeptide and nucleotide sequences, respectively. Column 3 shows the clone IDs of the Incyte clones in which nucleic acids encoding each PPHKP were identified, and column 4 shows the cDNA libraries from which these clones were isolated. Column 5 shows Incyte clones and their corresponding cDNA libraries. Clones for which cDNA libraries are not indicated were derived from pooled cDNA libraries. In some cases, GenBank sequence identifiers are also shown in column 5. The Incyte clones and GenBank cDNA sequences, where indicated, in column 5 were used to assemble the consensus nucleotide sequence of each PPHKP and are useful as fragments in hybridization technologies.

The columns of Table 2 show various properties of each of the polypeptides of the invention: column 1 references the SEQ ID NO; column 2 shows the number of amino acid residues in each polypeptide; column 3 shows potential phosphorylation sites; column 4 shows potential glycosylation sites; column 5 shows the amino acid residues comprising signature sequences and motifs; column 6 shows homologous sequences as identified by BLAST analysis; and column 7 shows analytical methods and in some cases, searchable databases to which the analytical methods were applied. The methods of column 7 were used to characterize each polypeptide through sequence homology and

protein motifs.

The columns of Table 3 show the tissue-specificity and diseases, disorders, or conditions associated with nucleotide sequences encoding PPHKP. The first column of Table 3 lists the nucleotide SEQ ID NOs. Column 2 lists fragments of the nucleotide sequences of column 1. These fragments are useful, for example, in hybridization or amplification technologies to identify SEQ ID NO:12-22 and to distinguish between SEQ ID NO:12-22 and related polynucleotide sequences. The polypeptides encoded by these fragments are useful, for example, as immunogenic peptides. Column 3 lists tissue categories which express PPHKP as a fraction of total tissues expressing PPHKP. Column 4 lists diseases, disorders, or conditions associated with those tissues expressing PPHKP as a fraction of total tissues expressing PPHKP. Column 5 lists the vectors used to subclone each cDNA library. Of particular note is the tissue specific expression of SEQ ID NO:14, SEQ ID NO:17, and SEQ ID NO:22. About 75% of the cDNA libraries expressing SEQ ID NO:14 are derived from nervous system tissue, particularly brain and spinal cord tissue. About 88% of the cDNA libraries expressing SEQ ID NO:17 are derived from nervous system tissue, particularly brain tissue. About 82% of the cDNA libraries expressing SEQ ID NO:22 are derived from gastrointestinal tissue.

The columns of Table 4 show descriptions of the tissues used to construct the cDNA libraries from which cDNA clones encoding PPHKP were isolated. Column 1 references the nucleotide SEQ ID NOs, column 2 shows the cDNA libraries from which these clones were isolated, and column 3 shows the tissue origins and other descriptive information relevant to the cDNA libraries in column 2.

SEQ ID NO:12 maps to chromosome 7 within the interval from 84.40 to 90.30 centiMorgans, to and to chromosome 11 within the interval from 89.80 to 90.70 centiMorgans. The interval on chromosome 7 from 84.40 to 90.30 centiMorgans also contains ESTs associated with B-cell CLL/lymphoma 7b. SEQ ID NO:16 maps to chromosome 20 within the interval from the p-terminus to 6.20 centiMorgans. This interval also contains ESTs associated with protein tyrosine phosphatase, non-receptor type substrate 1. SEQ ID NO:21 maps to chromosome 22 within the interval from 0.0 to 40.2 centiMorgans. This interval also contains ESTs associated with leukemia-associated phosphoprotein p18, a cytosolic phosphoprotein found in increased levels in the cells of various types of human acute leukemia.

The invention also encompasses PPHKP variants. A preferred PPHKP variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the PPHKP amino acid sequence, and which contains at least one functional or structural characteristic of PPHKP.

The invention also encompasses polynucleotides which encode PPHKP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:12-22, which encodes PPHKP. The polynucleotide

sequences of SEQ ID NO:12-22, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding PPHKP. In particular, such a variant polynucleotide sequence will have at least about 80%, or alternatively at least about 90%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding PPHKP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:12-22 which has at least about 80%, or alternatively at least about 90%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:12-22. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of PPHKP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding PPHKP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring PPHKP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode PPHKP and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring PPHKP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding PPHKP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding PPHKP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode PPHKP and PPHKP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding PPHKP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:12-22 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (PE Biosystems, Foster City CA), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (PE Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (PE Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding PPHKP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) *Nucleic Acids Res.* 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) *PCR Methods Applic.* 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) *Nucleic Acids Res.*

19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, PE Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode PPHKP may be cloned in recombinant DNA molecules that direct expression of PPHKP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express PPHKP.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter PPHKP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number

5,837,458; Chang, C.-C. et al. (1999) *Nat. Biotechnol.* 17:793-797; Christians, F.C. et al. (1999) *Nat. Biotechnol.* 17:259-264; and Crameri, A. et al. (1996) *Nat. Biotechnol.* 14:315-319) to alter or improve the biological properties of PPHKP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

15 In another embodiment, sequences encoding PPHKP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) *Nucleic Acids Symp. Ser.* 7:215-223; Horn, T. et al. (1980) *Nucleic Acids Symp. Ser.* 7:225-232.) Alternatively, PPHKP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) *Science* 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (PE Biosystems). Additionally, the amino acid sequence of PPHKP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

30 In order to express a biologically active PPHKP, the nucleotide sequences encoding PPHKP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding PPHKP. Such elements may vary in their strength and

specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding PPHKP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding PPHKP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used.

(See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding PPHKP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding PPHKP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Bitter, G.A. et al. (1987) *Methods Enzymol.* 153:516-544; Scorer, C.A. et al. (1994) *Bio/Technology* 12:181-184; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; and Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) *Cancer Gen. Ther.* 5(6):350-356; Yu, M. et al. (1993) *Proc. Natl. Acad. Sci.*

USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding PPHKP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding PPHKP can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding PPHKP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of PPHKP are needed, e.g. for the production of antibodies, vectors which direct high level expression of PPHKP may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of PPHKP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, supra; and Scorer, supra.)

Plant systems may also be used for expression of PPHKP. Transcription of sequences encoding PPHKP may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, supra; Broglie, supra; and Winter, supra.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding PPHKP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses PPHKP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma

virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of PPHKP in cell lines is preferred. For example, sequences encoding PPHKP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *apr* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech),  $\beta$  glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding PPHKP is inserted within a marker gene sequence, transformed cells containing sequences encoding PPHKP can be identified by the absence of marker gene function. Alternatively,

a marker gene can be placed in tandem with a sequence encoding PPHKP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding PPHKP and that express PPHKP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of PPHKP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on PPHKP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding PPHKP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding PPHKP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding PPHKP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode PPHKP may be designed to contain signal sequences which

direct secretion of PPHKP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation,

- 5 phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct  
10 modification and processing of the foreign protein.

- In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding PPHKP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric PPHKP protein containing a heterologous moiety that can be recognized by a commercially available antibody may  
15 facilitate the screening of peptide libraries for inhibitors of PPHKP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their  
20 cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the PPHKP encoding sequence and the heterologous protein  
25 sequence, so that PPHKP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

- In a further embodiment of the invention, synthesis of radiolabeled PPHKP may be achieved  
30 in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, <sup>35</sup>S-methionine.

- PPHKP of the present invention or fragments thereof may be used to screen for compounds  
35 that specifically bind to PPHKP. At least one and up to a plurality of test compounds may be

screened for specific binding to PPHKP. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of PPHKP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which PPHKP binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express PPHKP, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing PPHKP or cell membrane fractions which contain PPHKP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either PPHKP or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with PPHKP, either in solution or affixed to a solid support, and detecting the binding of PPHKP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

PPHKP of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of PPHKP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for PPHKP activity, wherein PPHKP is combined with at least one test compound, and the activity of PPHKP in the presence of a test compound is compared with the activity of PPHKP in the absence of the test compound. A change in the activity of PPHKP in the presence of the test compound is indicative of a compound that modulates the activity of PPHKP. Alternatively, a test compound is combined with an in vitro or cell-free system comprising PPHKP under conditions suitable for PPHKP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of PPHKP may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding PPHKP or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal

models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) *Science* 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) *Clin. Invest.* 97:1999-2002; Wagner, K.U. et al. (1997) *Nucleic Acids Res.* 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding PPHKP may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) *Science* 282:1145-1147).

Polynucleotides encoding PPHKP can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding PPHKP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress PPHKP, e.g., by secreting PPHKP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) *Biotechnol. Annu. Rev.* 4:55-74).

## THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of PPHKP and protein phosphatase and kinase proteins. In addition, the expression of PPHKP is closely associated with gastrointestinal, inflamed, nervous, proliferating and cancerous tissue. Therefore, PPHKP appears to play a role in gastrointestinal disorders, immune system disorders, neurological disorders, and cell proliferative disorders, including cancer. In the treatment of disorders associated with increased PPHKP expression or activity, it is desirable to decrease the expression or activity of PPHKP. In the treatment of disorders associated with decreased PPHKP

expression or activity, it is desirable to increase the expression or activity of PPHKP.

Therefore, in one embodiment, PPHKP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PPHKP. Examples of such disorders include, but are not limited to, a gastrointestinal disorder, such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha<sub>1</sub>-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; an immune system disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other

demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease; prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome; fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, 5 encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis; inherited, metabolic, 10 endocrine, and toxic myopathies; myasthenia gravis, periodic paralysis; mental disorders including mood, anxiety, and schizophrenic disorders; seasonal affective disorder (SAD); akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, 15 atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, 20 pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a vector capable of expressing PPHKP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PPHKP including, but not limited to, those described above.

25 In a further embodiment, a composition comprising a substantially purified PPHKP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PPHKP including, but not limited to, those provided above.

30 In still another embodiment, an agonist which modulates the activity of PPHKP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PPHKP including, but not limited to, those listed above.

35 In a further embodiment, an antagonist of PPHKP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PPHKP. Examples of such disorders include, but are not limited to, those gastrointestinal disorders, immune system disorders, neurological disorders, and cell proliferative disorders, including cancer described above. In one

aspect, an antibody which specifically binds PPHKP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express PPHKP.

In an additional embodiment, a vector expressing the complement of the polynucleotide  
5 encoding PPHKP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PPHKP including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made  
10 by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of PPHKP may be produced using methods which are generally known in the  
15 art. In particular, purified PPHKP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind PPHKP. Antibodies to PPHKP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit  
20 dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with PPHKP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral  
25 gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to PPHKP have an amino acid sequence consisting of at least about 5 amino acids, and generally will  
30 consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of PPHKP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to PPHKP may be prepared using any technique which provides for  
35 the production of antibody molecules by continuous cell lines in culture. These include, but are not

limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cote, R.J. et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:2026-2030; and Cole, S.P. et al. (1984) *Mol. Cell Biol.* 62:109-120.)

5 In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) *Proc. Natl. Acad. Sci. USA* 81:6851-6855; Neuberger, M.S. et al. (1984) *Nature* 312:604-608; and Takeda, S. et al. (1985) *Nature* 314:452-454.) Alternatively, techniques described for the production of single  
10 chain antibodies may be adapted, using methods known in the art, to produce PPHKP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) *Proc. Natl. Acad. Sci. USA* 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte  
15 population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:3833-3837; Winter, G. et al. (1991) *Nature* 349:293-299.)

Antibody fragments which contain specific binding sites for PPHKP may also be generated. For example, such fragments include, but are not limited to, F(ab')<sub>2</sub> fragments produced by pepsin  
20 digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) *Science* 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired  
25 specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between PPHKP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering PPHKP epitopes is generally used, but a competitive binding assay  
30 may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for PPHKP. Affinity is expressed as an association constant, K<sub>a</sub>, which is defined as the molar concentration of PPHKP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions.

35 The K<sub>a</sub> determined for a preparation of polyclonal antibodies, which are heterogeneous in their

affinities for multiple PPHKP epitopes, represents the average affinity, or avidity, of the antibodies for PPHKP. The  $K_d$  determined for a preparation of monoclonal antibodies, which are monospecific for a particular PPHKP epitope, represents a true measure of affinity. High-affinity antibody preparations with  $K_d$  ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the PPHKP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with  $K_d$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of PPHKP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of PPHKP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al., supra.)

In another embodiment of the invention, the polynucleotides encoding PPHKP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding PPHKP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding PPHKP. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et

al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding PPHKP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in PPHKP expression or regulation causes disease, the expression of PPHKP from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in PPHKP are treated by constructing mammalian expression vectors encoding PPHKP and introducing these vectors by mechanical means into PPHKP-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of PPHKP include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). PPHKP may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or  $\beta$ -actin genes), (ii) an inducible promoter (e.g., the

tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) *Proc. Natl. Acad. Sci. USA* 89:5547-5551; Gossen, M. et al. (1995) *Science* 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) *Curr. Opin. Biotechnol.* 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the  
5 FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, *supra*), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding PPHKP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver  
10 polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) *Virology* 52:456-467), or by electroporation (Neumann, E. et al. (1982) *EMBO J.* 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

15 In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to PPHKP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding PPHKP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences  
20 required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) *Proc. Natl. Acad. Sci. USA* 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al.  
25 (1987) *J. Virol.* 61:1647-1650; Bender, M.A. et al. (1987) *J. Virol.* 61:1639-1646; Adam, M.A. and A.D. Miller (1988) *J. Virol.* 62:3802-3806; Dull, T. et al. (1998) *J. Virol.* 72:8463-8471; Zufferey, R. et al. (1998) *J. Virol.* 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by  
30 reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4<sup>+</sup> T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) *J. Virol.* 71:7020-7029; Bauer, G. et al. (1997) *Blood* 89:2259-2267; Bonyhadi, M.L. (1997) *J. Virol.* 71:4707-4716; Ranga, U. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:1201-1206; Su, L. (1997) *Blood* 89:2283-  
35 2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding PPHKP to cells which have one or more genetic abnormalities with respect to the expression of PPHKP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) *Transplantation* 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) *Annu. Rev. Nutr.* 19:511-544; and Verma, I.M. and N. Somia (1997) *Nature* 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding PPHKP to target cells which have one or more genetic abnormalities with respect to the expression of PPHKP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing PPHKP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) *Exp. Eye Res.* 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) *J. Virol.* 73:519-532 and Xu, H. et al. (1994) *Dev. Biol.* 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding PPHKP to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) *Curr. Opin. Biotechnol.* 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full-length genomic RNA,

resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for PPHKP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of PPHKP-coding RNAs and the synthesis of high levels of PPHKP in vector transduced cells. While  
5 alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of PPHKP into a variety of cell types. The specific transduction of a subset of  
10 cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions  
15 -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E.  
20 and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme  
25 molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding PPHKP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA,  
30 GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

35 Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared

by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding PPHKP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding PPHKP. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased PPHKP expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding PPHKP may be therapeutically useful, and in the treatment of disorders associated with decreased PPHKP expression or activity, a compound which specifically promotes expression of the polynucleotide encoding PPHKP may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding PPHKP is exposed to at least one test compound thus obtained. The sample

may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding PPHKP are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding PPHKP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruce, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruce, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of PPHKP, antibodies to PPHKP, and mimetics, agonists, antagonists, or inhibitors of PPHKP.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary,

intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising PPHKP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, PPHKP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example PPHKP or fragments thereof, antibodies of PPHKP, and agonists, antagonists or inhibitors of PPHKP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the  $ED_{50}$  (the dose therapeutically effective in 50% of the population) or  $LD_{50}$  (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the  $LD_{50}/ED_{50}$  ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the  $ED_{50}$  with little or no toxicity.

The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu\text{g}$  to 100,000  $\mu\text{g}$ , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## DIAGNOSTICS

In another embodiment, antibodies which specifically bind PPHKP may be used for the diagnosis of disorders characterized by expression of PPHKP, or in assays to monitor patients being treated with PPHKP or agonists, antagonists, or inhibitors of PPHKP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for PPHKP include methods which utilize the antibody and a label to detect PPHKP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring PPHKP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of PPHKP expression. Normal or standard values for PPHKP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibody to PPHKP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of PPHKP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding PPHKP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences,

complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of PPHKP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of PPHKP, and to monitor regulation of PPHKP levels during therapeutic intervention.

5 In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding PPHKP or closely related molecules may be used to identify nucleic acid sequences which encode PPHKP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the  
10 probe identifies only naturally occurring sequences encoding PPHKP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the PPHKP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:12-22 or from  
15 genomic sequences including promoters, enhancers, and introns of the PPHKP gene.

Means for producing specific hybridization probes for DNAs encoding PPHKP include the cloning of polynucleotide sequences encoding PPHKP or PPHKP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA  
20 polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as  $^{32}\text{P}$  or  $^{35}\text{S}$ , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding PPHKP may be used for the diagnosis of disorders associated with expression of PPHKP. Examples of such disorders include, but are not limited to, a  
25 gastrointestinal disorder, such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis,  
30 passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha<sub>1</sub>-  
35 antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein

obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; an immune system disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease,

5 adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic

10 gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative

15 colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive

20 neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease; prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome; fatal familial insomnia, nutritional and metabolic diseases of the

25 nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis; inherited, metabolic,

30 endocrine, and toxic myopathies; myasthenia gravis, periodic paralysis; mental disorders including mood, anxiety, and schizophrenic disorders; seasonal affective disorder (SAD); akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and a cell proliferative disorder, such as actinic keratosis, arteriosclerosis,

35 atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis,

paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. The polynucleotide sequences encoding PPHKP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered PPHKP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding PPHKP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding PPHKP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding PPHKP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of PPHKP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding PPHKP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the

development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

5 Additional diagnostic uses for oligonucleotides designed from the sequences encoding PPHKP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding PPHKP, or a fragment of a polynucleotide complementary to the polynucleotide encoding PPHKP, and will be employed under optimized conditions for identification of a specific gene or  
10 condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding PPHKP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic  
15 disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding PPHKP are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause  
20 differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP (isSNP), are capable of identifying polymorphisms by comparing the  
25 sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

30 Methods which may also be used to quantify the expression of PPHKP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of  
35 interest is presented in various dilutions and a spectrophotometric or colorimetric response gives

rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described in Seilhamer, J.J. et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, incorporated herein by reference. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, antibodies specific for PPHKP, or PPHKP or fragments thereof may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed

molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) *Mol. Carcinog.* 24:153-159; Steiner, S. and N.L. Anderson (2000) *Toxicol. Lett.* 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical

density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for PPHKP to quantify the levels of PPHKP expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111; Mendozze, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are

incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding PPHKP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, e.g., Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding PPHKP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps.

Often the placement of a gene on the chromosome of another mammalian species, such as mouse,

may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, PPHKP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between PPHKP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with PPHKP, or fragments thereof, and washed. Bound PPHKP is then detected by methods well known in the art. Purified PPHKP can also be coated directly onto plates for use in the aforementioned drug screening techniques.

Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding PPHKP specifically compete with a test compound for binding PPHKP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with PPHKP.

In additional embodiments, the nucleotide sequences which encode PPHKP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. 60/154,141, are hereby expressly incorporated by reference.

## EXAMPLES

### I. Construction of cDNA Libraries

RNA was purchased from Clontech or isolated from tissues described in Table 4. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A+) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), pcDNA2.1 plasmid (Invitrogen, Carlsbad CA), or pINCY plasmid (Incyte Genomics, Palo Alto CA). Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or ElectroMAX DH10B from Life Technologies.

### II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1

ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) *Anal. Biochem.* 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

### III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows.

Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (PE Biosystems) thermal cyclers or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (PE Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (PE Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, *supra*, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VI.

The polynucleotide sequences derived from cDNA sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 5 summarizes the tools, programs, and algorithms used and provides applicable descriptions, references, and threshold parameters. The first column of Table 5 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the greater the homology between two sequences). Sequences were analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments were generated using the default parameters specified by the clustal algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned

sequences.

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and PFAM to acquire annotation using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA.

The full length polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GenBank databases (described above), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and Hidden Markov Model (HMM)-based protein family databases such as PFAM. HMM is a probabilistic approach which analyzes consensus primary structures of gene families. (See, e.g., Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.)

The programs described above for the assembly and analysis of full length polynucleotide and amino acid sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:12-22. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies were described in The Invention section above.

#### IV. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel, 1995, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum \{length(Seq. 1), length(Seq. 2)\}}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the

product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

The results of northern analyses are reported as a percentage distribution of libraries in which the transcript encoding PPHKP occurred. Analysis involved the categorization of cDNA libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease/condition categories included cancer, inflammation, trauma, cell proliferation, neurological, and pooled. For each category, the number of libraries expressing the sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease- or condition-specific expression are reported in Table 3.

#### **V. Chromosomal Mapping of PPHKP Encoding Polynucleotides**

The cDNA sequences which were used to assemble SEQ ID NO:12-22 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:12-22 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 5). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO., to that map location.

The genetic map locations of SEQ ID NO:12, SEQ ID NO:16 and SEQ ID NO:21 are described in The Invention as ranges, or intervals, of human chromosomes. More than one map location is reported for SEQ ID NO:12, indicating that previously mapped sequences having similarity, but not complete identity, to SEQ ID NO:12 were assembled into their respective clusters. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination

frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation-hybrid markers whose sequences were included in each of the clusters. Diseases associated with the public and Incyte sequences located within the indicated intervals are also reported in the Invention where applicable.

#### VI. Extension of PPHKP Encoding Polynucleotides

The full length nucleic acid sequences of SEQ ID NO:12-22 were produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing  $Mg^{2+}$ ,  $(NH_4)_2SO_4$ , and  $\beta$ -mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100  $\mu$ l PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5  $\mu$ l of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5  $\mu$ l to 10  $\mu$ l aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (PE Biosystems).

In like manner, the polynucleotide sequences of SEQ ID NO:12-22 are used to obtain 5' regulatory sequences using the procedure above, along with oligonucleotides designed for such extension, and an appropriate genomic library.

## VII. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:12-22 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 µCi of [ $\gamma$ -<sup>32</sup>P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10<sup>7</sup> counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon

membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

### VIII. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Skena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Skena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

#### Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)<sup>+</sup> RNA is purified using the oligo-(dT) cellulose method. Each poly(A)<sup>+</sup> RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)<sup>+</sup> RNA with

GEMBRIGHT kits (Incyte). Specific control poly(A)<sup>+</sup> RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37 °C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85 °C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 µl 5X SSC/0.2% SDS.

#### 10 Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 µg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 µl of the array element DNA, at an average concentration of 100 ng/µl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60 °C followed by washes in 0.2% SDS and distilled water as before.

#### 30 Hybridization

Hybridization reactions contain 9 µl of sample mixture consisting of 0.2 µg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65 °C for 5 minutes and is aliquoted onto the microarray surface and covered

with an 1.8 cm<sup>2</sup> coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 µl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash  
5 buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

#### Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines  
10 at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

15 In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is  
20 typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that  
25 location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

30 The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and  
35 measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping

emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

#### IX. Complementary Polynucleotides

Sequences complementary to the PPHKP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring PPHKP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of PPHKP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the PPHKP-encoding transcript.

#### X. Expression of PPHKP

Expression and purification of PPHKP is achieved using bacterial or virus-based expression systems. For expression of PPHKP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac (tac)* hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express PPHKP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of PPHKP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding PPHKP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, PPHKP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step,

affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from PPHKP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified PPHKP obtained by these methods can be used directly in the assays shown in Examples XI and XV.

#### **XI. Demonstration of PPHKP Activity**

Protein kinase activity is measured by quantifying the phosphorylation of a protein substrate by PPHKP in the presence of gamma-labeled  $^{32}\text{P}$ -ATP. PPHKP is incubated with the protein substrate,  $^{32}\text{P}$ -ATP, and an appropriate kinase buffer. The  $^{32}\text{P}$  incorporated into the substrate is separated from free  $^{32}\text{P}$ -ATP by electrophoresis and the incorporated  $^{32}\text{P}$  is counted using a radioisotope counter. The amount of incorporated  $^{32}\text{P}$  is proportional to the activity of PPHKP. A determination of the specific amino acid residue phosphorylated is made by phosphoamino acid analysis of the hydrolyzed protein.

Alternatively, protein phosphatase activity is measured by the hydrolysis of P-nitrophenyl phosphate (PNPP). PPHKP is incubated with PNPP in HEPES buffer pH 7.5, in the presence of 0.1% b-mercaptoethanol at 37°C for 60 min. The reaction is stopped by the addition of 6 ml of 10 N NaOH and the increase in light absorbance of the reaction mixture at 410 nm resulting from the hydrolysis of PNPP is measured using a spectrophotometer. The increase in light absorbance is proportional to the activity of PPHKP (Diamond R.H. et al (1994) Mol Cell Biol 14:3752-62).

#### **XII. Functional Assays**

PPHKP function is assessed by expressing the sequences encoding PPHKP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT plasmid (Life Technologies) and pCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10  $\mu\text{g}$  of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2  $\mu\text{g}$  of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP;

Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of PPHKP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding PPHKP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding PPHKP and other genes of interest can be analyzed by northern analysis or microarray techniques.

### **XIII. Production of PPHKP Specific Antibodies**

PPHKP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the PPHKP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (PE Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-PPHKP activity by, for example, binding the peptide or PPHKP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

#### **XIV. Purification of Naturally Occurring PPHKP Using Specific Antibodies**

Naturally occurring or recombinant PPHKP is substantially purified by immunoaffinity chromatography using antibodies specific for PPHKP. An immunoaffinity column is constructed by covalently coupling anti-PPHKP antibody to an activated chromatographic resin, such as

5 CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing PPHKP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of PPHKP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt  
10 antibody/PPHKP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and PPHKP is collected.

#### **XV. Identification of Molecules Which Interact with PPHKP**

PPHKP, or biologically active fragments thereof, are labeled with <sup>125</sup>I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules  
15 previously arrayed in the wells of a multi-well plate are incubated with the labeled PPHKP, washed, and any wells with labeled PPHKP complex are assayed. Data obtained using different concentrations of PPHKP are used to calculate values for the number, affinity, and association of PPHKP with the candidate molecules.

Alternatively, molecules interacting with PPHKP are analyzed using the yeast two-hybrid  
20 system as described in Fields, S. and O. Song (1989, Nature 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

PPHKP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S.  
25 Patent No. 6,057,101).

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it  
30 should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

TABLE 1

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
1	12	365665	SYNORAT01	365665H1 (SYNORAT01), 1751084F6 (LIVRTUT01), 1753091T6 (LIVRTUT01), 2649915H1 (THYMFET02), 2741857F6 (BRSTTUT14), 2914093F6 (KIDNTUT15), 3129019T6 (LUNGFTUT12), 3187065H1 (THYMNON04), 3875544H1 (HEARNOT06)
2	13	760934	BRAITUT02	405845R6 (EOSIHET02), 732719F1 (LUNGNOT03), 760934H1 (BRAITUT02), 760934R7 (BRAITUT02)
3	14	926043	BRAINOT04	926043H1 (BRAINOT04), 998403X309D1 (KIDNTUT01), 1835276R6 (BRAINON01), 3268349H1 (BRAINOT20), SXLA0113V1, g921351, g4454847
4	15	1234795	LUNGFET03	1234795F1 (LUNGFET03), 1234795H1 (LUNGFET03), 3091866H1 (BRSTNOT19), SXBA00067V1, SXBA01121V1, SXBA01178V1, SXBA00158V1, SXBA01170V1, SXBA01184V1
5	16	1271505	TESTTUT02	1271505H1 (TESTTUT02), 1271505X309F1 (TESTTUT02), 1395158X11 (THYRNOT03), 1432510R1 (BEPINON01), 1518249F6 (BLADTUT04), 1829986H1 (THP1AZT01), 2236286F6 (PANCUTUT02), 3590149F6 (293TF5T01)
6	17	1385073	BRAITUT08	533781F1 (BRAINOT03), 533781R1 (BRAINOT03), 927395R6 (BRAINOT04), 1286637H1 (BRAINOT11), 1287320F6 (BRAINOT11), 1385073F6 (BRAITUT08), 1385073H1 (BRAITUT08), 3334151F6 (BRAIFET01), 4339301H1 (BRAUNOT02), 4754877H1 (BRAHNOT01), SANA02528F1
7	18	1606974	LUNGNOT15	542579R6 (OVARNOT02), 938242R1 (CERVNOT01), 1472528T1 (LUNGNOT03), 1606974H1 (LUNGNOT15), 1729937H1 (BRSTTUT08), g2207365
8	19	1829744	THP1AZT01	817088R6 (OVARTUT01), 1757328R6 (PITUNOT03), 1829744H1 (THP1AZT01), 1899355T6 (BLADTUT06), 3818445F6 (BONSTUT01)
9	20	4030831	BRAINOT23	601077R1 (BRSTNOT02), 863312T1 (BRAITUT03), 1333294F1 (COLNNOT13), 1623843F6 (BRAITUT13), 1657143F6 (URETUT01), 1863418F6 (PROSNOT19), 1889555F6 (BLADTUT07), 3692741H1 (HEAANOT01), 4030831H1 (BRAINOT23), 4691493H1 (BRAENOT02), SARA02402F1, SARA02217F1
10	21	5039718	COLHTUT01	286342F1 (EOSIHET02), 398033R6 (PITUNOT02), 854522R1 (NGANNOT01), 1632128H1 (COLNNOT19), 1711324F6 (PROSNOT16), 2620936X309D4 (KERANOT02), 2941327H1 (HEAONOT03), 4895491H1 (LIVRTUT12), 5039718H2 (COLHTUT01), SCHA00577V1, SCJA02058V1
11	22	5595281	COLCDIT03	2211301F6 (SINTFET03), 2211301T6 (SINTFET03), 5595281H1 (COLCDIT03)

TABLE 2

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Homologous Sequence	Analytical Methods
1	329	S26 T110 T164 S103 T123 T207	N3 N7 N36 N227	Signal Peptide: M1 - G24 STK/CaM kinase-like region: H138 - L318	myosin heavy chain kinase B (g1903458) [Dictyostelium discoideum]	BLAST-GenBank, BLAST-Prodcom, SPScan, Motifs
2	141	T33		Mammalian acid phosphatase: I20 - E141	lysophosphatidic acid phosphatase (g6691475) [Homo sapiens]	BLAST-GenBank, BLAST - DOMO, Motifs
3	447	T22 S28 S63 S109 T121 S183 S190 T226 S283 S331 T369 T7 T114 S163 T204 S242 S246 S272 S292 T303 S381 T412	N11 N33 N273 N347	Protein phosphatase 2A regulatory subunit PR55 signature: E79 - N93 Protein phosphatase 2A regulatory subunit PR55 domains: T22 - D68, L86 - R126, T146 - G348, K389 - K441 Region of protein phosphatase 2A regulatory subunit PR55: M1 - F365	protein phosphatase 2A subunit beta (g190426) [Homo sapiens]	BLAST-GenBank, BLAST-DOMO, BLAST-Prodcom, BLIMPS-BLOCKS, BLIMPS-PRINTS, Motifs
4	666	T280 S628 S21 S91 S214 S266 S369 S421 S422 T434 T590 S613 S618 S643 T12 T23 S72 S82 S393 T439 S468 S471 T479 T528 S597 S605 S610	N38 N49 N190 N212 N300 N641	Tyrosine specific protein phosphatases active site: L220 - T280 Y phosphatase signature: N237 - R278 VH1-type dual specificity phosphatase: V85 - I298	neuronal tyrosine threonine phosphatase 1 (g1781037) [Mus musculus]	BLAST-GenBank, BLAST-DOMO, BLAST-Prodcom, HMMER - PFAM, BLIMPS-PRINTS, ProfileScan, Motifs

TABLE 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Homologous Sequence	Analytical Methods
5	358	T54 T90 T140 T210 S215 S14 T27 T95 S232 S238		Protein kinase signature: G142 - L315 Tyrosine kinase catalytic domain signature: A242 - S264 Protein kinase domain: E78 - R303	kinase (g4827159) [Rattus sp.]	BLAST-GenBank, BLAST-DOMO, HMMER - PFAM, BLIMPS-PRINTS, Motifs
6	470	T7 T91 T103 S118 T276 S341 T439 S452 S464 T21 T46 T51 T103 S138 S292 T391 S451 Y135 Y460		Protein kinase signature: Y24 - I286 Tyrosine kinase catalytic domain signature: Y135 - Y153, V201 - E223 Protein kinase domain: D25 - T276	calmodulin-binding protein (g349075) [Rattus norvegicus]	BLAST-GenBank, BLAST-DOMO, HMMER - PFAM, BLIMPS-PRINTS, Motifs
7	150	S41 T43	N7	Signal Peptide: M1 - Y28 Tyrosine specific protein phosphatases active site: V93 - M105 Y phosphatase signature: P64 - A107 Tyrosine specific protein phosphatase domains: V93 - G103, R126 - A141 I74 - I125	COS41.7 (g1764017) [Ciona intestinalis] similar to tyrosine phosphatase	BLAST-GenBank, HMMER - PFAM, BLIMPS-PRINTS, BLIMPS-BLOCKS, SpScan, Motifs
8	253	T81 S250 T8 S34 T121 S128 T216 S29 T132 S228 S229		Protein kinases signatures and profile: L158 - L170 Protein kinase domain: L33 - S187 Protein hydrolase domain: G42 - R247	YGR262c putative protein kinase (g1536954) [Saccharomyces cerevisiae]	BLAST-GenBank, BLAST-Prodom, HMMER - PFAM, Motifs

TABLE 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Homologous Sequence	Analytical Methods
9	442	S84 S190 S215 T231 S241 S339 S428 T4 S255 T355		Non-receptor tyrosine kinase domains: E64 - C229, G246 - F331, V399 - D436	non-receptor tyrosine kinase (g974334) [Dictyostelium discoideum]	BLAST-GenBank, BLAST-Prodome, Motifs
10	659	S37 S70 S132 S140 S266 S286 T292 S484 S542 S552 S564 S568 S569 S649 S653 T65 S254 T395 T638 Y369	N614	Signal Peptide: M1 - V19 VH1-type dual specificity phosphatase: I332 - S470	MAP kinase phosphatase (g6714641) [Drosophila melanogaster]	BLAST-GenBank, BLAST-DOMO, SpScan, Motifs
11	145	T32 S33 S55 S141 T35 S41 T89 S91 S104 T110 S27 S45 S51 T58			17-kDa PKC-potentiated inhibitory protein of PP1 (g2570154) [Sus scrofa]	BLAST-GenBank, Motifs

TABLE 3

SEQ ID NO:	Unique Fragments	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
12	164-223 758-817	Cardiovascular (0.286) Reproductive (0.286) Hematopoietic/Immune (0.143)	Cancer (0.524) Inflammation/trauma (0.334) Cell Proliferation (0.095)	PSPORT1
13	158-217	Cardiovascular (0.200) Reproductive (0.200) Hematopoietic/Immune (0.200) Nervous (0.200)	Cancer (0.800) Inflammation (0.300)	PSPORT1
14	1-60 1189-1248	Nervous (0.750) Reproductive (0.083) Urologic (0.083) Developmental (0.083)	Cancer (0.333) Inflammation/trauma (0.333) Neurological (0.250)	PSPORT1
15	163-222 1081-1140	Gastrointestinal (0.385) Cardiovascular (0.154) Nervous (0.154) Reproductive (0.154)	Cancer (0.692) Inflammation/trauma (0.308) Cell Proliferation (0.231)	pINCY
16	326-385	Reproductive (0.288) Gastrointestinal (0.212) Hematopoietic/Immune (0.192)	Cancer (0.577) Inflammation/trauma (0.327) Cell Proliferation (0.308)	pINCY
17	326-385 1298-1357	Nervous (0.885) Developmental (0.077)	Inflammation/trauma (0.384) Neurological (0.231) Cancer (0.192)	pINCY
18	251-310	Reproductive (0.372) Cardiovascular (0.140) Endocrine (0.093)	Cancer (0.558) Inflammation/trauma (0.233) Cell Proliferation (0.209)	pINCY
19	380-439	Reproductive (0.368) Cardiovascular (0.158) Gastrointestinal (0.105) Nervous (0.105) Hematopoietic/Immune (0.105)	Cancer (0.474) Inflammation/trauma (0.369) Cell Proliferation (0.211)	pINCY
20	213-272 813-872	Nervous (0.304) Reproductive (0.278) Gastrointestinal (0.104)	Cancer (0.435) Inflammation/trauma (0.357) Cell Proliferation (0.096)	pINCY
21	219-279 921-980	Reproductive (0.343) Gastrointestinal (0.194) Hematopoietic/Immune (0.134)	Cancer (0.552) Inflammation/trauma (0.314) Cell Proliferation (0.090)	pINCY
22	233-292	Gastrointestinal (0.818) Developmental (0.091) Reproductive (0.091)	Inflammation/trauma (0.636) Cancer (0.182) Cell Proliferation (0.091)	pINCY

TABLE 4

Nucleotide SEQ ID NO:	Library	Library Description
12	SYNORAT01	The library was constructed using RNA isolated from synovial membrane tissue removed from the elbow of a 51-year-old Asian female with rheumatoid arthritis.
13	BRAITUT02	The library was constructed using RNA isolated from brain tumor tissue removed from the frontal lobe of a 58-year-old Caucasian male during excision of a cerebral meningeal lesion. Pathology indicated a grade 2 metastatic hypernephroma. Patient history included a grade 2 renal cell carcinoma, insomnia, and chronic airway obstruction. Family history included a malignant neoplasm of the kidney.
14	BRAINOT04	The library was constructed using RNA isolated from the brain tissue of a 44-year-old Caucasian male with a cerebral hemorrhage. The tissue, which contained coagulated blood, came from the choroid plexus of the right anterior temporal lobe. Family history included coronary artery disease and myocardial infarction.
15	LUNGFET03	The library was constructed using RNA isolated from lung tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation.
16	TESTTUT02	The library was constructed using RNA isolated from testicular tumor removed from a 31-year-old Caucasian male during unilateral orchiectomy. Pathology indicated embryonal carcinoma.
17	BRAITUT08	The library was constructed using RNA isolated from brain tumor tissue removed from the left frontal lobe of a 47-year-old Caucasian male during excision of cerebral meningeal tissue. Pathology indicated grade 4 fibrillary astrocytoma with focal tumoral radionecrosis. Patient history included cerebrovascular disease, deficiency anemia, hyperlipidemia, epilepsy, and tobacco use. Family history included cerebrovascular disease and a malignant prostate neoplasm.
18	LUNGNOT15	The library was constructed using RNA isolated from lung tissue removed from a 69-year-old Caucasian male during a segmental lung resection. Pathology for the associated tumor tissue indicated residual grade 3 invasive squamous cell carcinoma. Patient history included acute myocardial infarction, prostatic hyperplasia, and malignant skin neoplasm. Family history included cerebrovascular disease, type I diabetes, acute myocardial infarction, and arteriosclerotic coronary disease.

TABLE 4 (cont.)

Nucleotide SEQ ID NO:	Library	Library Description
19	THP1AZT01	The library was constructed using RNA isolated from THP-1 promonocyte cells treated for three days with 0.8 micromolar 5-aza-2'-deoxycytidine. THP-1 (ATCC TIB 202) is a human promonocyte line derived from peripheral blood of a 1-year-old Caucasian male with acute monocytic leukemia (Int. J. Cancer (1980) 26:171).
20	BRAINT023	The library was constructed using RNA isolated from right temporal lobe tissue removed from a 45-year-old Black male during a brain lobectomy. Pathology for the associated tumor tissue indicated dysembryoplastic neuroepithelial tumor of the right temporal lobe. The right temporal region dura was consistent with calcifying pseudo-tumor of the neuraxis. The patient presented with convulsive intractable epilepsy, partial epilepsy, and memory disturbance. Patient history included obesity, meningitis, backache, unspecified sleep apnea, acute stress reaction, acquired knee deformity, and chronic sinusitis. Family history included obesity, benign hypertension, cirrhosis of the liver, alcohol abuse, hyperlipidemia, cerebrovascular disease, and type II diabetes.
21	COLHTUT01	The library was constructed using RNA isolated from colon tumor tissue removed from the hepatic flexure of a 55-year-old Caucasian male during right hemicolectomy, incidental appendectomy, and permanent colostomy. Pathology indicated invasive grade 3 adenocarcinoma. Patient history included benign hypertension, anxiety, abnormal blood chemistry, blepharitis, heart block, osteoporosis, acne, and hyperplasia of the prostate. Family history included prostate cancer, acute myocardial infarction, stroke, and atherosclerotic coronary artery disease.
22	COLCDIT03	The library was constructed using RNA isolated from diseased colon polyp tissue removed from the cecum of a 67-year-old female. Pathology indicated a benign cecum polyp. Pathology for the associated tumor tissue indicated invasive grade 3 adenocarcinoma that arose in tubulovillous adenoma forming a fungating mass in the cecum.

Table 5

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	PE Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	PE Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	PE Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value= 1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.06E-6 Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger; and, if applicable, Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322.	Score=10-50 bits for PFAM hits, depending on individual protein families

Table 5 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score > GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

- 5           a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11,  
            b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11,  
            c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and  
10           d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11.

2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-11.

15

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

20           5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:12-22.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

25

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

30

9. A method for producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

35

b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

11. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

- 5           a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22,  
            b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22,  
            c) a polynucleotide sequence complementary to a),  
            d) a polynucleotide sequence complementary to b), and  
10           e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.

13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

- a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization  
20           complex is formed between said probe and said target polynucleotide or fragments thereof, and  
            b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

- a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and  
30           b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A composition comprising an effective amount of a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

17. A composition of claim 16, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-11.

18. A method for treating a disease or condition associated with decreased expression of functional PPHKP, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

10       a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

          b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

15

21. A method for treating a disease or condition associated with decreased expression of functional PPHKP, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

20       a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

          b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

25

24. A method for treating a disease or condition associated with overexpression of functional PPHKP, comprising administering to a patient in need of such treatment a composition of claim 23.

30

25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:

          a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

35       b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying

a compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

5 a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

10 c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

20 c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

28. A method for assessing toxicity of a test compound, said method comprising:

a) treating a biological sample containing nucleic acids with the test compound;

25 b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;

30 c) quantifying the amount of hybridization complex; and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

## SEQUENCE LISTING

<110> INCYTE GENOMICS, INC.  
 YUE, Henry  
 TANG, Y. Tom  
 BANDMAN, Olga  
 HILLMAN, Jennifer L.  
 BAUGHN, Mariah R.  
 AZIMZAI, Valda  
 LU, Dyung Aina M.

<120> PROTEIN PHOSPHATASE AND KINASE PROTEINS

<130> PF-0742 PCT

<140> To Be Assigned

<141> Herewith

<150> 60/154,141

<151> 1999-09-15

<160> 22

<170> PERL Program

<210> 1

<211> 329

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID NO: 365665CD1

<400> 1

Met	Leu	Asn	Cys	Ser	Gln	Asn	Ser	Ser	Ser	Ser	Ser	Val	Trp	Trp
1				5					10					15
Leu	Lys	Ser	Pro	Ala	Phe	Ser	Ser	Gly	Ser	Ser	Glu	Gly	Asp	Ser
				20					25					30
Pro	Trp	Ser	Tyr	Leu	Asn	Ser	Ser	Gly	Ser	Ser	Trp	Val	Ser	Leu
				35					40					45
Pro	Gly	Lys	Met	Arg	Lys	Glu	Ile	Leu	Glu	Ala	Arg	Thr	Leu	Gln
				50					55					60
Pro	Asp	Asp	Phe	Glu	Lys	Leu	Leu	Ala	Gly	Val	Arg	His	Asp	Trp
				65					70					75
Leu	Phe	Gln	Arg	Leu	Glu	Asn	Thr	Gly	Val	Phe	Lys	Pro	Ser	Gln
				80					85					90
Leu	His	Arg	Ala	His	Ser	Ala	Leu	Leu	Leu	Lys	Tyr	Ser	Lys	Lys
				95					100					105
Ser	Glu	Leu	Trp	Thr	Ala	Gln	Glu	Thr	Ile	Val	Tyr	Leu	Gly	Asp
				110					115					120
Tyr	Leu	Thr	Val	Lys	Lys	Lys	Gly	Arg	Gln	Arg	Asn	Ala	Phe	Trp
				125					130					135
Val	His	His	Leu	His	Gln	Glu	Glu	Ile	Leu	Gly	Arg	Tyr	Val	Gly
				140					145					150
Lys	Asp	Tyr	Lys	Glu	Gln	Lys	Gly	Leu	Trp	His	His	Phe	Thr	Asp
				155					160					165
Val	Glu	Arg	Gln	Met	Thr	Ala	Gln	His	Tyr	Val	Thr	Glu	Phe	Asn
				170					175					180
Lys	Arg	Leu	Tyr	Glu	Gln	Asn	Ile	Pro	Thr	Gln	Ile	Phe	Tyr	Ile

Pro Ser Thr Ile	185	Leu Leu Ile Leu Glu Asp Lys Thr Ile Lys Gly	190	195
	200		205	210
Cys Ile Ser Val	215	Glu Pro Tyr Ile Leu Gly Glu Phe Val Lys Leu	220	225
Ser Asn Asn Thr	230	Lys Val Val Lys Thr Glu Tyr Lys Ala Thr Glu	235	240
Tyr Gly Leu Ala	245	Tyr Gly His Phe Ser Tyr Glu Phe Ser Asn His	250	255
Arg Asp Val Val	260	Val Asp Leu Gln Gly Trp Val Thr Gly Asn Gly	265	270
Lys Gly Leu Ile	275	Tyr Leu Thr Asp Pro Gln Ile His Ser Val Asp	280	285
Gln Lys Val Phe	290	Thr Thr Asn Phe Gly Lys Arg Gly Ile Phe Tyr	295	300
Phe Phe Asn Asn	305	Gln His Val Glu Cys Asn Glu Ile Cys His Arg	310	315
Leu Ser Leu Thr	320	Arg Pro Ser Met Glu Lys Pro Cys Lys Ser	325	

&lt;210&gt; 2

&lt;211&gt; 141

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID NO: 760934CD1

&lt;400&gt; 2

Met Leu Phe Thr Phe His Arg Glu Ser Leu Gln Met Ala Val Gly	1	5	10	15
Pro Phe Leu His Ile Leu Glu Ser Asn Leu Leu Lys Ala Met Asp	20	25	30	
Ser Ala Thr Ala Pro Asp Lys Ile Arg Lys Leu Tyr Leu Tyr Ala	35	40	45	
Ala His Asp Val Thr Phe Ile Pro Leu Leu Met Thr Leu Gly Ile	50	55	60	
Phe Asp His Lys Trp Pro Pro Phe Ala Val Asp Leu Thr Met Glu	65	70	75	
Leu Tyr Gln His Leu Glu Ser Lys Glu Trp Phe Val Gln Leu Tyr	80	85	90	
Tyr His Gly Lys Glu Gln Val Pro Arg Gly Cys Pro Asp Gly Leu	95	100	105	
Cys Pro Leu Asp Met Phe Leu Asn Ala Met Ser Val Tyr Thr Leu	110	115	120	
Ser Pro Glu Lys Tyr His Ala Leu Cys Ser Gln Thr Gln Val Met	125	130	135	
Glu Val Gly Asn Glu Glu	140			

&lt;210&gt; 3

&lt;211&gt; 447

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID NO: 926043CD1

&lt;400&gt; 3

```

Met Gly Glu Asp Thr Asp Thr Arg Lys Ile Asn His Ser Phe Leu
 1          5          10          15
Arg Asp His Ser Tyr Val Thr Glu Ala Asp Ile Phe Ser Thr Val
          20          25          30
Glu Phe Asn His Thr Gly Glu Leu Leu Ala Thr Gly Asp Lys Gly
          35          40          45
Gly Arg Val Val Ile Phe Gln Arg Glu Pro Glu Ser Lys Asn Ala
          50          55          60
Pro His Ser Gln Gly Glu Tyr Asp Val Tyr Ser Thr Phe Gln Ser
          65          70          75
His Glu Pro Glu Phe Asp Tyr Leu Lys Ser Leu Glu Ile Glu Glu
          80          85          90
Lys Ile Asn Lys Ile Lys Trp Leu Pro Gln Gln Asn Ala Ala His
          95          100          105
Ser Leu Leu Ser Thr Asn Asp Lys Thr Ile Lys Leu Trp Lys Ile
          110          115          120
Thr Glu Arg Asp Lys Arg Pro Glu Gly Tyr Asn Leu Lys Asp Glu
          125          130          135
Glu Gly Lys Leu Lys Asp Leu Ser Thr Val Thr Ser Leu Gln Val
          140          145          150
Pro Val Leu Lys Pro Met Asp Leu Met Val Glu Val Ser Pro Arg
          155          160          165
Arg Ile Phe Ala Asn Gly His Thr Tyr His Ile Asn Ser Ile Ser
          170          175          180
Val Asn Ser Asp Cys Glu Thr Tyr Met Ser Ala Asp Asp Leu Arg
          185          190          195
Ile Asn Leu Trp His Leu Ala Ile Thr Asp Arg Ser Phe Asn Ile
          200          205          210
Val Asp Ile Lys Pro Ala Asn Met Glu Asp Leu Thr Glu Val Ile
          215          220          225
Thr Ala Ser Glu Phe His Pro His His Cys Asn Leu Phe Val Tyr
          230          235          240
Ser Ser Ser Lys Gly Ser Leu Arg Leu Cys Asp Met Arg Ala Ala
          245          250          255
Ala Leu Cys Asp Lys His Ser Lys Leu Phe Glu Glu Pro Glu Asp
          260          265          270
Pro Ser Asn Arg Ser Phe Phe Ser Glu Ile Ile Ser Ser Val Ser
          275          280          285
Asp Val Lys Phe Ser His Ser Gly Arg Tyr Met Leu Thr Arg Asp
          290          295          300
Tyr Leu Thr Val Lys Val Trp Asp Leu Asn Met Glu Ala Arg Pro
          305          310          315
Ile Glu Thr Tyr Gln Val His Asp Tyr Leu Arg Ser Lys Leu Cys
          320          325          330
Ser Leu Tyr Glu Asn Asp Cys Ile Phe Asp Lys Phe Glu Cys Ala
          335          340          345
Trp Asn Gly Ser Asp Ser Val Ile Met Thr Gly Ala Tyr Asn Asn
          350          355          360
Phe Phe Arg Met Phe Asp Arg Asn Thr Lys Arg Asp Val Thr Leu
          365          370          375
Glu Ala Ser Arg Glu Ser Ser Lys Pro Arg Ala Val Leu Lys Pro
          380          385          390
Arg Arg Val Cys Val Gly Gly Lys Arg Arg Arg Asp Asp Ile Ser
          395          400          405
Val Asp Ser Leu Asp Phe Thr Lys Lys Ile Leu His Thr Ala Trp
          410          415          420
His Pro Ala Glu Asn Ile Ile Ala Ile Ala Ala Thr Asn Asn Leu
          425          430          435
Tyr Ile Phe Gln Asp Lys Val Asn Ser Asp Met His

```

440

445

<210> 4  
 <211> 666  
 <212> PRT  
 <213> Homo sapiens  
 <220>  
 <221> misc\_feature  
 <223> Incyte ID NO: 1234795CD1

<400> 4  
 Met Ala His Glu Met Ile Gly Thr Gln Ile Val Thr Glu Arg Leu  
 1 5 10 15  
 Val Ala Leu Leu Glu Ser Gly Thr Glu Lys Val Leu Leu Ile Asp  
 20 25 30  
 Ser Arg Pro Phe Val Glu Tyr Asn Thr Ser His Ile Leu Glu Ala  
 35 40 45  
 Ile Asn Ile Asn Cys Ser Lys Leu Met Lys Arg Arg Leu Gln Gln  
 50 55 60  
 Asp Lys Val Leu Ile Thr Glu Leu Ile Gln His Ser Ala Lys His  
 65 70 75  
 Lys Val Asp Ile Asp Cys Ser Gln Lys Val Val Tyr Asp Gln  
 80 85 90  
 Ser Ser Gln Asp Val Ala Ser Leu Ser Ser Asp Cys Phe Leu Thr  
 95 100 105  
 Val Leu Leu Gly Lys Leu Glu Lys Ser Phe Asn Ser Val His Leu  
 110 115 120  
 Leu Ala Gly Gly Phe Ala Glu Phe Ser Arg Cys Phe Pro Gly Leu  
 125 130 135  
 Cys Glu Gly Lys Ser Thr Leu Val Pro Thr Cys Ile Ser Gln Pro  
 140 145 150  
 Cys Leu Pro Val Ala Asn Ile Gly Pro Thr Arg Ile Leu Pro Asn  
 155 160 165  
 Leu Tyr Leu Gly Cys Gln Arg Asp Val Leu Asn Lys Glu Leu Met  
 170 175 180  
 Gln Gln Asn Gly Ile Gly Tyr Val Leu Asn Ala Ser Asn Thr Cys  
 185 190 195  
 Pro Lys Pro Asp Phe Ile Pro Glu Ser His Phe Leu Arg Val Pro  
 200 205 210  
 Val Asn Asp Ser Phe Cys Glu Lys Ile Leu Pro Trp Leu Asp Lys  
 215 220 225  
 Ser Val Asp Phe Ile Glu Lys Ala Lys Ala Ser Asn Gly Cys Val  
 230 235 240  
 Leu Val His Cys Leu Ala Gly Ile Ser Arg Ser Ala Thr Ile Ala  
 245 250 255  
 Ile Ala Tyr Ile Met Lys Arg Met Asp Met Ser Leu Asp Glu Ala  
 260 265 270  
 Tyr Arg Phe Val Lys Glu Lys Arg Pro Thr Ile Ser Pro Asn Phe  
 275 280 285  
 Asn Phe Leu Gly Gln Leu Leu Asp Tyr Glu Lys Lys Ile Lys Asn  
 290 295 300  
 Gln Thr Gly Ala Ser Gly Pro Lys Ser Lys Leu Lys Leu Leu His  
 305 310 315  
 Leu Glu Lys Pro Asn Glu Pro Val Pro Ala Val Ser Glu Gly Gly  
 320 325 330  
 Gln Lys Ser Glu Thr Pro Leu Ser Pro Pro Cys Ala Asp Ser Ala  
 335 340 345  
 Thr Ser Glu Ala Ala Gly Gln Arg Pro Val His Pro Ala Ser Val  
 350 355 360

Pro	Ser	Val	Pro	Ser	Val	Gln	Pro	Ser	Leu	Leu	Glu	Asp	Ser	Pro	
			365						370					375	
Leu	Val	Gln	Ala	Leu	Ser	Gly	Leu	His	Leu	Ser	Ala	Asp	Arg	Leu	
			380						385					390	
Glu	Asp	Ser	Asn	Lys	Leu	Lys	Arg	Ser	Phe	Ser	Leu	Asp	Ile	Lys	
			395						400					405	
Ser	Val	Ser	Tyr	Ser	Ala	Ser	Met	Ala	Ala	Ser	Leu	His	Gly	Phe	
			410						415					420	
Ser	Ser	Ser	Glu	Asp	Ala	Leu	Glu	Tyr	Tyr	Lys	Pro	Ser	Thr	Thr	
			425						430					435	
Leu	Asp	Gly	Thr	Asn	Lys	Leu	Cys	Gln	Phe	Ser	Pro	Val	Gln	Glu	
			440						445					450	
Leu	Ser	Glu	Gln	Thr	Pro	Glu	Thr	Ser	Ser	Leu	Ile	Arg	Arg	Lys	
			455						460					465	
Pro	Ala	Ser	Pro	Arg	Ser	Cys	Arg	Pro	Pro	Gly	Leu	Gln	Thr	Ala	
			470						475					480	
Arg	Ala	Ser	Asp	Cys	Ile	Arg	Ser	Glu	Pro	Ala	Ala	Val	Ala	Pro	
			485						490					495	
Pro	Arg	Gly	Pro	Phe	Tyr	Leu	His	Cys	Ile	Glu	Val	Gly	Ala	Trp	
			500						505					510	
Arg	Thr	Ile	Thr	Thr	Pro	Ala	Ser	Phe	Ser	Ala	Phe	Pro	Pro	Ala	
			515						520					525	
Ser	Ser	Thr	Ser	Arg	Ser	Leu	Leu	Ala	Trp	Ala	Leu	Lys	Gly	Trp	
			530						535					540	
His	Ser	Asp	Ile	Leu	Ala	Pro	Gln	Thr	Ser	Thr	Pro	Ser	Leu	Thr	
			545						550					555	
Ser	Ser	Trp	Tyr	Phe	Ala	Thr	Glu	Ser	Ser	His	Phe	Tyr	Ser	Ala	
			560						565					570	
Ser	Ala	Ile	Tyr	Gly	Gly	Ser	Ala	Ser	Tyr	Ser	Ala	Tyr	Ser	Cys	
			575						580					585	
Ser	Gln	Leu	Pro	Thr	Cys	Gly	Asp	Gln	Val	Tyr	Ser	Val	Arg	Arg	
			590						595					600	
Arg	Gln	Lys	Pro	Ser	Asp	Arg	Ala	Asp	Ser	Arg	Arg	Ser	Trp	His	
			605						610					615	
Glu	Glu	Ser	Pro	Phe	Glu	Lys	Gln	Phe	Lys	Arg	Arg	Ser	Cys	Gln	
			620						625					630	
Met	Glu	Phe	Gly	Glu	Ser	Ile	Met	Ser	Glu	Asn	Arg	Ser	Arg	Glu	
			635						640					645	
Glu	Leu	Gly	Lys	Val	Gly	Ser	Gln	Ser	Ser	Phe	Ser	Gly	Ser	Met	
			650						655					660	
Glu	Ile	Ile	Glu	Val	Ser										
			665												

<210> 5  
 <211> 358  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID NO: 1271505CD1

<400> 5  
 Met Arg Ala Thr Pro Leu Ala Ala Pro Ala Gly Ser Leu Ser Arg  
 1 5 10 15  
 Lys Lys Arg Leu Glu Leu Asp Asp Asn Leu Asp Thr Glu Arg Pro  
 20 25 30  
 Val Gln Lys Arg Ala Arg Ser Gly Pro Gln Pro Arg Leu Pro Pro  
 35 40 45  
 Cys Leu Leu Pro Leu Ser Pro Pro Thr Ala Pro Asp Arg Ala Thr

Ala Val Ala Thr	Ala Ser Arg Leu Gly	Pro Tyr Val Leu Leu Glu	50	55	60
			65	70	75
Pro Glu Glu Gly	Gly Arg Ala Tyr Gln	Ala Leu His Cys Pro Thr	80	85	90
Gly Thr Glu Tyr	Thr Cys Lys Val Tyr	Pro Val Gln Glu Ala Leu	95	100	105
Ala Val Leu Glu	Pro Tyr Ala Arg Leu	Pro Pro His Lys His Val	110	115	120
Ala Arg Pro Thr	Glu Val Leu Ala Gly	Thr Gln Leu Leu Tyr Ala	125	130	135
Phe Phe Thr Arg	Thr His Gly Asp Met	His Ser Leu Val Arg Thr	140	145	150
Arg His Arg Ile	Pro Glu Pro Glu Ala	Ala Val Leu Phe Arg Gln	155	160	165
Met Ala Thr Ala	Leu Ala His Cys His	Gln His Gly Leu Val Leu	170	175	180
Arg Asp Leu Lys	Leu Cys Arg Phe Val	Phe Ala Asp Arg Glu Arg	185	190	195
Lys Lys Leu Val	Leu Glu Asn Leu Glu	Asp Ser Cys Val Leu Thr	200	205	210
Gly Pro Asp Asp	Ser Leu Trp Asp Lys	His Ala Cys Pro Ala Tyr	215	220	225
Val Gly Pro Glu	Ile Leu Ser Ser Arg	Ala Ser Tyr Ser Gly Lys	230	235	240
Ala Ala Asp Val	Trp Ser Leu Gly Val	Ala Leu Phe Thr Met Leu	245	250	255
Ala Gly His Tyr	Pro Phe Gln Asp Ser	Glu Pro Val Leu Leu Phe	260	265	270
Gly Lys Ile Arg	Arg Gly Ala Tyr Ala	Leu Pro Ala Gly Leu Ser	275	280	285
Ala Pro Ala Arg	Cys Leu Val Arg Cys	Leu Leu Arg Arg Glu Pro	290	295	300
Ala Glu Arg Leu	Thr Ala Thr Gly Ile	Leu Leu His Pro Trp Leu	305	310	315
Arg Gln Asp Pro	Met Pro Leu Ala Pro	Thr Arg Ser His Leu Trp	320	325	330
Glu Ala Ala Gln	Val Val Pro Asp Gly	Leu Gly Leu Asp Glu Ala	335	340	345
Arg Glu Glu Glu	Gly Asp Arg Glu Val	Val Leu Tyr Gly	350	355	

&lt;210&gt; 6

&lt;211&gt; 470

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID NO: 1385073CD1

&lt;400&gt; 6

Met Pro Phe Gly Cys	Val Thr Leu Gly Asp	Lys Lys Asn Tyr Asn	1	5	10	15
Gln Pro Ser Glu Val	Thr Asp Arg Tyr Asp	Leu Gly Gln Val Ile	20	25	30	
Lys Thr Glu Glu Phe	Cys Glu Ile Phe Arg	Ala Lys Asp Lys Thr	35	40	45	
Thr Gly Lys Leu His	Thr Cys Lys Lys Phe	Gln Lys Arg Asp Gly	50	55	60	

Arg	Lys	Val	Arg	Lys	Ala	Ala	Lys	Asn	Glu	Ile	Gly	Ile	Leu	Lys	65	70	75
Met	Val	Lys	His	Pro	Asn	Ile	Leu	Gln	Leu	Val	Asp	Val	Phe	Val	80	85	90
Thr	Arg	Lys	Glu	Tyr	Phe	Ile	Phe	Leu	Glu	Leu	Ala	Thr	Gly	Arg	95	100	105
Glu	Val	Phe	Asp	Trp	Ile	Leu	Asp	Gln	Gly	Tyr	Tyr	Ser	Glu	Arg	110	115	120
Asp	Thr	Ser	Asn	Val	Val	Arg	Gln	Val	Leu	Glu	Ala	Val	Ala	Tyr	125	130	135
Leu	His	Ser	Leu	Lys	Ile	Val	His	Arg	Asn	Leu	Lys	Leu	Glu	Asn	140	145	150
Leu	Val	Tyr	Tyr	Asn	Arg	Leu	Lys	Asn	Ser	Lys	Ile	Val	Ile	Ser	155	160	165
Asp	Phe	His	Leu	Ala	Lys	Leu	Glu	Asn	Gly	Leu	Ile	Lys	Glu	Pro	170	175	180
Cys	Gly	Thr	Pro	Glu	Tyr	Leu	Ala	Pro	Glu	Val	Val	Gly	Arg	Gln	185	190	195
Arg	Tyr	Gly	Arg	Pro	Val	Asp	Cys	Trp	Ala	Ile	Gly	Val	Ile	Met	200	205	210
Tyr	Ile	Leu	Leu	Ser	Gly	Asn	Pro	Pro	Phe	Tyr	Glu	Glu	Val	Glu	215	220	225
Glu	Asp	Asp	Tyr	Glu	Asn	His	Asp	Lys	Asn	Leu	Phe	Arg	Lys	Ile	230	235	240
Leu	Ala	Gly	Asp	Tyr	Glu	Phe	Asp	Ser	Pro	Tyr	Trp	Asp	Asp	Ile	245	250	255
Ser	Gln	Ala	Ala	Lys	Asp	Leu	Val	Thr	Arg	Leu	Met	Glu	Val	Glu	260	265	270
Gln	Asp	Gln	Arg	Ile	Thr	Ala	Glu	Glu	Ala	Ile	Ser	His	Glu	Trp	275	280	285
Ile	Ser	Gly	Asn	Ala	Ala	Ser	Asp	Lys	Asn	Ile	Lys	Asp	Gly	Val	290	295	300
Cys	Ala	Gln	Ile	Glu	Lys	Asn	Phe	Ala	Arg	Ala	Lys	Trp	Lys	Lys	305	310	315
Ala	Val	Arg	Val	Thr	Thr	Leu	Met	Lys	Arg	Leu	Arg	Ala	Pro	Glu	320	325	330
Gln	Ser	Ser	Thr	Ala	Ala	Ala	Gln	Ser	Ala	Ser	Ala	Thr	Asp	Thr	335	340	345
Ala	Thr	Pro	Gly	Ala	Ala	Asp	Arg	Ser	Ala	Thr	Pro	Ala	Thr	Asp	350	355	360
Gly	Ser	Ala	Thr	Pro	Ala	Thr	Asp	Gly	Ser	Val	Thr	Pro	Ala	Thr	365	370	375
Asp	Gly	Ser	Ile	Thr	Pro	Ala	Thr	Asp	Gly	Ser	Val	Thr	Pro	Ala	380	385	390
Thr	Asp	Arg	Ser	Ala	Thr	Pro	Ala	Thr	Asp	Gly	Arg	Ala	Thr	Pro	395	400	405
Ala	Thr	Glu	Glu	Ser	Thr	Val	Pro	Thr	Thr	Gln	Ser	Ser	Ala	Met	410	415	420
Leu	Ala	Thr	Lys	Ala	Ala	Ala	Thr	Pro	Glu	Pro	Ala	Met	Ala	Gln	425	430	435
Pro	Asp	Ser	Thr	Ala	Pro	Glu	Gly	Ala	Thr	Gly	Gln	Ala	Pro	Pro	440	445	450
Ser	Ser	Lys	Gly	Glu	Glu	Ala	Ala	Gly	Tyr	Ala	Gln	Glu	Ser	Gln	455	460	465
Arg	Glu	Glu	Ala	Ser											470		

<210> 7  
 <211> 150  
 <212> PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID NO: 1606974CD1

&lt;400&gt; 7

Met	Gly	Val	Gln	Pro	Pro	Asn	Phe	Ser	Trp	Val	Leu	Pro	Gly	Arg	1	5	10	15
Leu	Ala	Gly	Leu	Ala	Leu	Pro	Arg	Leu	Pro	Ala	His	Tyr	Gln	Phe	20	25	30	
Leu	Leu	Asp	Leu	Gly	Val	Arg	His	Leu	Val	Ser	Leu	Thr	Glu	Arg	35	40	45	
Gly	Pro	Pro	His	Ser	Asp	Ser	Cys	Pro	Gly	Leu	Thr	Leu	His	Arg	50	55	60	
Leu	Arg	Ile	Pro	Asp	Phe	Cys	Pro	Pro	Ala	Pro	Asp	Gln	Ile	Asp	65	70	75	
Arg	Phe	Val	Gln	Ile	Val	Asp	Glu	Ala	Asn	Ala	Arg	Gly	Glu	Ala	80	85	90	
Val	Gly	Val	His	Cys	Ala	Leu	Gly	Phe	Gly	Arg	Thr	Gly	Thr	Met	95	100	105	
Leu	Ala	Cys	Tyr	Leu	Val	Lys	Glu	Arg	Gly	Leu	Ala	Ala	Gly	Asp	110	115	120	
Ala	Ile	Ala	Glu	Ile	Arg	Arg	Leu	Arg	Pro	Gly	Ser	Ile	Glu	Thr	125	130	135	
Tyr	Glu	Gln	Glu	Lys	Ala	Val	Phe	Gln	Phe	Tyr	Gln	Arg	Thr	Lys	140	145	150	

&lt;210&gt; 8

&lt;211&gt; 253

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID NO: 1829744CD1

&lt;400&gt; 8

Met	Ala	Ala	Ala	Arg	Ala	Thr	Thr	Pro	Ala	Asp	Gly	Glu	Glu	Pro	1	5	10	15
Ala	Pro	Glu	Ala	Glu	Ala	Leu	Ala	Ala	Ala	Arg	Glu	Arg	Ser	Ser	20	25	30	
Arg	Phe	Leu	Ser	Gly	Leu	Glu	Leu	Val	Lys	Gln	Gly	Ala	Glu	Ala	35	40	45	
Arg	Val	Phe	Arg	Gly	Arg	Phe	Gln	Gly	Arg	Ala	Ala	Val	Ile	Lys	50	55	60	
His	Arg	Phe	Pro	Lys	Gly	Tyr	Arg	His	Pro	Ala	Leu	Glu	Ala	Arg	65	70	75	
Leu	Gly	Arg	Arg	Arg	Thr	Val	Gln	Glu	Ala	Arg	Ala	Leu	Leu	Arg	80	85	90	
Cys	Arg	Arg	Ala	Gly	Ile	Ser	Ala	Pro	Val	Val	Phe	Phe	Val	Asp	95	100	105	
Tyr	Ala	Ser	Asn	Cys	Leu	Tyr	Met	Glu	Glu	Ile	Glu	Gly	Ser	Val	110	115	120	
Thr	Val	Arg	Asp	Tyr	Ile	Gln	Ser	Thr	Met	Glu	Thr	Glu	Lys	Thr	125	130	135	
Pro	Gln	Gly	Leu	Ser	Asn	Leu	Ala	Lys	Thr	Ile	Gly	Gln	Val	Leu	140	145	150	
Ala	Arg	Met	His	Asp	Glu	Asp	Leu	Ile	His	Gly	Asp	Leu	Thr	Thr	155	160	165	

Ser	Asn	Met	Leu	Leu	Lys	Pro	Pro	Leu	Glu	Gln	Leu	Asn	Ile	Val	
				170					175					180	
Leu	Ile	Asp	Phe	Gly	Leu	Ser	Phe	Ile	Ser	Ala	Leu	Pro	Glu	Asp	
				185					190					195	
Lys	Gly	Val	Asp	Leu	Tyr	Val	Leu	Glu	Lys	Ala	Phe	Leu	Ser	Thr	
				200					205					210	
His	Pro	Asn	Thr	Glu	Thr	Val	Phe	Glu	Ala	Phe	Leu	Lys	Ser	Tyr	
				215					220					225	
Ser	Thr	Ser	Ser	Lys	Lys	Ala	Arg	Pro	Val	Leu	Lys	Lys	Leu	Asp	
				230					235					240	
Glu	Val	Arg	Leu	Arg	Gly	Arg	Lys	Arg	Ser	Met	Val	Gly			
				245					250						

&lt;210&gt; 9

&lt;211&gt; 442

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID NO: 4030831CD1

&lt;400&gt; 9

Met	Arg	Arg	Thr	Arg	Arg	Pro	Arg	Phe	Val	Leu	Met	Asn	Lys	Met	
1				5					10					15	
Asp	Asp	Leu	Asn	Leu	His	Tyr	Arg	Phe	Leu	Asn	Trp	Arg	Arg	Arg	
				20					25					30	
Ile	Arg	Glu	Ile	Arg	Glu	Val	Arg	Ala	Phe	Arg	Tyr	Gln	Glu	Arg	
				35					40					45	
Phe	Lys	His	Ile	Leu	Val	Asp	Gly	Asp	Thr	Leu	Ser	Tyr	His	Gly	
				50					55					60	
Asn	Ser	Gly	Glu	Val	Gly	Cys	Tyr	Val	Ala	Ser	Arg	Pro	Leu	Thr	
				65					70					75	
Lys	Asp	Ser	Asn	Tyr	Phe	Glu	Val	Ser	Ile	Val	Asp	Ser	Gly	Val	
				80					85					90	
Arg	Gly	Thr	Ile	Ala	Val	Gly	Leu	Val	Pro	Gln	Tyr	Tyr	Ser	Leu	
				95					100					105	
Asp	His	Gln	Pro	Gly	Trp	Leu	Pro	Asp	Ser	Val	Ala	Tyr	His	Ala	
				110					115					120	
Asp	Asp	Gly	Lys	Leu	Tyr	Asn	Gly	Arg	Ala	Lys	Gly	Arg	Gln	Phe	
				125					130					135	
Gly	Ser	Lys	Cys	Asn	Ser	Gly	Asp	Arg	Ile	Gly	Cys	Gly	Ile	Glu	
				140					145					150	
Pro	Val	Ser	Phe	Asp	Val	Gln	Thr	Ala	Gln	Ile	Phe	Phe	Thr	Lys	
				155					160					165	
Asn	Gly	Lys	Arg	Val	Gly	Ser	Thr	Ile	Met	Pro	Met	Ser	Pro	Asp	
				170					175					180	
Gly	Leu	Phe	Pro	Ala	Val	Gly	Met	His	Ser	Leu	Gly	Glu	Glu	Val	
				185					190					195	
Arg	Leu	His	Leu	Asn	Ala	Glu	Leu	Gly	Arg	Glu	Asp	Asp	Ser	Val	
				200					205					210	
Met	Met	Val	Asp	Ser	Tyr	Glu	Asp	Glu	Trp	Gly	Arg	Leu	His	Asp	
				215					220					225	
Val	Arg	Val	Cys	Gly	Thr	Leu	Leu	Glu	Tyr	Leu	Gly	Lys	Gly	Lys	
				230					235					240	
Ser	Ile	Val	Asp	Val	Gly	Leu	Ala	Gln	Ala	Arg	His	Pro	Leu	Ser	
				245					250					255	
Thr	Arg	Ser	His	Tyr	Phe	Glu	Val	Glu	Ile	Val	Asp	Pro	Gly	Glu	
				260					265					270	
Lys	Cys	Tyr	Ile	Ala	Leu	Gly	Leu	Ala	Arg	Lys	Asp	Tyr	Pro	Lys	

	275		280		285
Asn Arg His Pro Gly Trp Ser Arg Gly Ser Val Ala Tyr His Ala					
	290		295		300
Asp Asp Gly Lys Ile Phe His Gly Ser Gly Val Gly Asp Pro Phe					
	305		310		315
Gly Pro Arg Cys Tyr Lys Gly Asp Ile Met Gly Cys Gly Ile Met					
	320		325		330
Phe Pro Arg Asp Tyr Ile Leu Asp Ser Glu Gly Asp Ser Asp Asp					
	335		340		345
Ser Cys Asp Thr Val Ile Leu Ser Pro Thr Ala Arg Ala Val Arg					
	350		355		360
Asn Val Arg Asn Val Met Tyr Leu His Gln Glu Gly Glu Glu Glu					
	365		370		375
Glu Glu Glu Glu Glu Glu Glu Glu Asp Gly Glu Glu Ile Glu Pro					
	380		385		390
Glu His Glu Gly Arg Lys Val Val Val Phe Phe Thr Arg Asn Gly					
	395		400		405
Lys Ile Ile Gly Lys Lys Asp Ala Val Val Pro Ser Gly Gly Phe					
	410		415		420
Phe Pro Thr Ile Gly Met Leu Ser Cys Gly Glu Lys Val Lys Val					
	425		430		435
Asp Leu His Pro Leu Ser Gly					
	440				

&lt;210&gt; 10

&lt;211&gt; 659

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID NO: 5039718CD1

&lt;400&gt; 10

Met Ala Leu Val Thr Val Ser Arg Ser Pro Pro Gly Ser Gly Ala		
1 5 10 15		
Ser Thr Pro Val Gly Pro Trp Asp Gln Ala Val Gln Arg Arg Ser		
20 25 30		
Arg Leu Gln Arg Arg Gln Ser Phe Ala Val Leu Arg Gly Ala Val		
35 40 45		
Leu Gly Leu Gln Asp Gly Gly Asp Asn Asp Asp Ala Ala Glu Ala		
50 55 60		
Ser Ser Glu Pro Thr Glu Lys Ala Pro Ser Glu Glu Glu Leu His		
65 70 75		
Gly Asp Gln Thr Asp Phe Gly Gln Gly Ser Gln Ser Pro Gln Lys		
80 85 90		
Gln Glu Glu Gln Arg Gln His Leu His Leu Met Val Gln Leu Leu		
95 100 105		
Arg Pro Gln Asp Asp Ile Arg Leu Ala Ala Gln Leu Glu Ala Pro		
110 115 120		
Arg Pro Pro Arg Leu Arg Tyr Leu Leu Val Val Ser Thr Arg Glu		
125 130 135		
Gly Glu Gly Leu Ser Gln Asp Glu Thr Val Leu Leu Gly Val Asp		
140 145 150		
Phe Pro Asp Ser Ser Ser Pro Ser Cys Thr Leu Gly Leu Val Leu		
155 160 165		
Pro Leu Trp Ser Asp Thr Gln Val Tyr Leu Asp Gly Asp Gly Gly		
170 175 180		
Phe Ser Val Thr Ser Gly Gly Gln Ser Arg Ile Phe Lys Pro Ile		
185 190 195		

Ser	Ile	Gln	Thr	Met	Trp	Ala	Thr	Leu	Gln	Val	Leu	His	Gln	Ala
				200					205					210
Cys	Glu	Ala	Ala	Leu	Gly	Ser	Gly	Leu	Val	Pro	Gly	Gly	Ser	Ala
				215					220					225
Leu	Thr	Trp	Ala	Ser	His	Tyr	Gln	Glu	Arg	Leu	Asn	Ser	Glu	Gln
				230					235					240
Ser	Cys	Leu	Asn	Glu	Trp	Thr	Ala	Met	Ala	Asp	Leu	Glu	Ser	Leu
				245					250					255
Arg	Pro	Pro	Ser	Ala	Glu	Pro	Gly	Gly	Ser	Ser	Glu	Gln	Glu	Gln
				260					265					270
Met	Glu	Gln	Ala	Ile	Arg	Ala	Glu	Leu	Trp	Lys	Val	Leu	Asp	Val
				275					280					285
Ser	Asp	Leu	Glu	Ser	Val	Thr	Ser	Lys	Glu	Ile	Arg	Gln	Ala	Leu
				290					295					300
Glu	Leu	Arg	Leu	Gly	Leu	Pro	Leu	Gln	Gln	Tyr	Arg	Asp	Phe	Ile
				305					310					315
Asp	Asn	Gln	Met	Leu	Leu	Val	Ala	Gln	Arg	Asp	Arg	Ala	Ser	Ser
				320					325					330
Arg	Ile	Phe	Pro	His	Leu	Tyr	Leu	Gly	Ser	Glu	Trp	Asn	Ala	Ala
				335					340					345
Asn	Leu	Glu	Glu	Leu	Gln	Arg	Asn	Arg	Val	Thr	His	Ile	Leu	Asn
				350					355					360
Met	Ala	Arg	Glu	Ile	Asp	Asn	Phe	Tyr	Pro	Glu	Arg	Phe	Thr	Tyr
				365					370					375
His	Asn	Val	Arg	Leu	Trp	Asp	Glu	Glu	Ser	Ala	Gln	Leu	Leu	Pro
				380					385					390
His	Trp	Lys	Glu	Thr	His	Arg	Phe	Ile	Glu	Ala	Ala	Arg	Ala	Gln
				395					400					405
Gly	Thr	His	Val	Leu	Val	His	Cys	Lys	Met	Gly	Val	Ser	Arg	Ser
				410					415					420
Ala	Ala	Thr	Val	Leu	Ala	Tyr	Ala	Met	Lys	Gln	Tyr	Glu	Cys	Ser
				425					430					435
Leu	Glu	Gln	Ala	Leu	Arg	His	Val	Gln	Glu	Leu	Arg	Pro	Ile	Ala
				440					445					450
Arg	Pro	Asn	Pro	Gly	Phe	Leu	Arg	Gln	Leu	Gln	Ile	Tyr	Gln	Gly
				455					460					465
Ile	Leu	Thr	Ala	Ser	Arg	Gln	Ser	His	Val	Trp	Glu	Gln	Lys	Val
				470					475					480
Gly	Gly	Val	Ser	Pro	Glu	Glu	His	Pro	Ala	Pro	Glu	Val	Ser	Thr
				485					490					495
Pro	Phe	Pro	Pro	Leu	Pro	Pro	Glu	Pro	Glu	Gly	Gly	Gly	Glu	Glu
				500					505					510
Lys	Val	Val	Gly	Met	Glu	Glu	Ser	Gln	Ala	Ala	Pro	Lys	Glu	Glu
				515					520					525
Pro	Gly	Pro	Arg	Pro	Arg	Ile	Asn	Leu	Arg	Gly	Val	Met	Arg	Ser
				530					535					540
Ile	Ser	Leu	Leu	Glu	Pro	Ser	Leu	Glu	Leu	Glu	Ser	Thr	Ser	Glu
				545					550					555
Thr	Ser	Asp	Met	Pro	Glu	Val	Phe	Ser	Ser	His	Glu	Ser	Ser	His
				560					565					570
Glu	Glu	Pro	Leu	Gln	Pro	Phe	Pro	Gln	Leu	Ala	Arg	Thr	Lys	Gly
				575					580					585
Gly	Gln	Gln	Val	Asp	Arg	Gly	Pro	Gln	Pro	Ala	Leu	Lys	Ser	Arg
				590					595					600
Gln	Ser	Val	Val	Thr	Leu	Gln	Gly	Ser	Ala	Val	Val	Ala	Asn	Arg
				605					610					615
Thr	Gln	Ala	Phe	Gln	Glu	Gln	Glu	Gln	Gly	Gln	Gly	Gln	Gly	Gln
				620					625					630
Gly	Glu	Pro	Cys	Ile	Ser	Ser	Thr	Pro	Arg	Phe	Arg	Lys	Val	Val
				635					640					645
Arg	Gln	Ala	Ser	Val	His	Asp	Ser	Gly	Glu	Glu	Gly	Glu	Ala	
				650					655					

<210> 11  
 <211> 145  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID NO: 5595281CD1

<400> 11  
 Met Leu Ser Ser Ser Pro Ala Ser Cys Thr Ser Pro Ser Pro Asp  
   1                  5                  10                  15  
 Gly Glu Asn Pro Cys Lys Lys Val His Trp Ala Ser Gly Arg Arg  
                   20                  25                  30  
 Arg Thr Ser Ser Thr Asp Ser Glu Ser Lys Ser His Pro Asp Ser  
                   35                  40                  45  
 Ser Lys Ile Pro Arg Ser Arg Arg Pro Ser Arg Leu Thr Val Lys  
                   50                  55                  60  
 Tyr Asp Arg Gly Gln Leu Gln Arg Trp Leu Glu Met Glu Gln Trp  
                   65                  70                  75  
 Val Asp Ala Gln Val Gln Glu Leu Phe Gln Asp Gln Ala Thr Pro  
                   80                  85                  90  
 Ser Glu Pro Glu Ile Asp Leu Glu Ala Leu Met Asp Leu Ser Thr  
                   95                  100                 105  
 Glu Glu Gln Lys Thr Gln Leu Glu Ala Ile Leu Gly Asn Cys Pro  
                  110                 115                 120  
 Arg Pro Thr Glu Ala Phe Ile Ser Glu Leu Leu Ser Gln Leu Lys  
                  125                 130                 135  
 Lys Leu Arg Arg Leu Ser Arg Pro Gln Lys  
                  140                 145

<210> 12  
 <211> 1884  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID NO: 365665CB1

<400> 12  
 ctgcactacc acagaggaag gaaatcagcc tggaaacatg ctaaactgca gccagaactc 60  
 cagctcatcc tcagtgtggt ggctgaaatc acctgcattt tccagtgggt cttctgaggg 120  
 ggacagccct tggctctatc tgaattccag tgggagttct tgggtttcat tgccgggaaa 180  
 gatgaggaaa gagatccttg aggctcgcac cttgcaacct gatgactttg aaaagctgtt 240  
 ggcaggagtg aggcattgatt ggctgtttca gagactagag aatacggggg tttttaagcc 300  
 cagtcaactc caccgagcac atagtgtctt tttgttaaaa tattcaaaaa aatctgaact 360  
 gtggacggcc caggaaacta ttgtctatct gggggactac ttgactgtga agaaaaaagg 420  
 cagacaaaaga aatgcttttt ggggtcatca tcttcatcaa gaagaaattc tggggaggta 480  
 tggttgggaaa gactataagg agcagaaggg gctctggcac cacttcactg atgtggagcg 540  
 acagatgacc gcacagcact atgtgacaga atttaacaag agactctatg aacaaaacat 600  
 tcccacccag atattctaca tcccatccac aatactactg atttttagagg acaagacaat 660  
 aaagggatgt atcagtgtgg agccttacat actggggagaa tttgtaaaat tgtcaaataa 720  
 cacgaaagtg gtgaaaacag aatacaaaagc cacagaatat ggcttggcct atggccattt 780  
 ttcttatgag ttttctaata atagagatgt tgtggtcgat ttacaagggt gggtaaccgg 840  
 taatggaaaa ggactcatct acctcacaga tccccagatt cactccgttg atcagaaagt 900  
 tttcactacc aattttggaa agagaggaat tttttacttc ttttaataacc agcatgtgga 960  
 atgtaatgaa atctgccatc gtcttttctt gactagacct tcaatggaga aaccatgtaa 1020

```

gtcataggct gtatggattg gtaatgtgac agaccttagt gatgctctca gatatctggt 1080
gatgctcgca aatatctggt tctctccttc caggcacata gaatacggca cagtctggtc 1140
ctttggggct tgggcagggc cgtgacacag gttctggcca atgatttgca agaggaattg 1200
atcagtatca ctttaagtcc tgcatttaat tggcagcaca agatcctgca gagcctcttt 1260
ccctctgcca cagttatcaa gaatgggtca ggagaccgct gcttctgggc ataagtcctg 1320
caaggaaagc aacatggaaa acagcccaa ctcacccatg agggatgaaa agcactcttg 1380
agaaaggcat gtgttgttta agccattgag attttagagc tttttgtcac tatctgtcaa 1440
gactgatact actggggcctt ttcctattga tttgggagtt ctttacatat taaaaaaatg 1500
tgagcctttg tgatacgaat tcaatttggt ttcctgtctt ttgacatttg actttgcata 1560
aaagtttatc tgtgcataat tttatatgta gttgaattca tcaatctttt attttgtatg 1620
gctttttggt tatgtataat acttagatcc tccttatact ctgagtttct ttctttttaa 1680
ttctcctgta tttccttcta gtataattaa atctgtaaaa agtaagatgg aagagtggta 1740
cagttttctt tatccagtct gtccttgatg ggcatttagg tagactggat aaagaaaatg 1800
tgatacatat acaccatgga acactatgtg tattaatcca ctctcacact gctatgaaga 1860
gatacctgag actgggtaat ttag                                     1884

```

&lt;210&gt; 13

&lt;211&gt; 784

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID NO: 760934CB1

&lt;400&gt; 13

```

cggacggtgg gcggacgcgt gggcggacgc gtgggaggac gcgtgggctc agccccctccc 60
atctcagttg tgttcacaac aggtctgtgt gctgggtgct gtgttggttg gagtcttcat 120
tcacagagga cccaagacc ggactcgggt ccatttctct ccttctggcc cgttgctggc 180
ccttatgcat tacttgctt tttctcagct ctgtctggac tcatgaggtc ctggggggcag 240
aatcagtcctg tgtcccctga atgctgagtg ccctctaggt gcttaaagt tgttacttt 300
tcacagggaa agtcttcaga tggcagtagg cccattcctc cacatcctag agagcaacct 360
gctgaaagcc atggactctg ccactgcccc cgacaagatc agaaagctgt atctctatgc 420
ggctcatgat gtgaccttca taccgctctt aatgaccctg gggatttttg accacaaatg 480
gccaccgttt gctgttgacc tgaccatgga actttaccag cacctggaat ctaaggagtg 540
gtttgtgcag ctctattacc acgggaagga gcaggtgccg agaggttgcc ctgatgggct 600
ctgcccgtct gacatgttct tgaatgccat gtcagtttat accttaagcc cagaaaaata 660
ccacgcactc tgctctcaaa ctcaggtgat ggaagttgga aatgaagagt aactgattta 720
taaaagcagg atgtgttgat tttaaaataa agtgccttta tacaatgaaa aaaaaaaaaa 780
aaaa                                              784

```

&lt;210&gt; 14

&lt;211&gt; 1657

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID NO: 926043CB1

&lt;400&gt; 14

```

ccagagcacc gggcacggcc ttcaatgggc gaggacacgg acacgcggaa aattaaccac 60
agcttctctgc gggaccacag ctatgtgact gaagctgaca tcttctctac cgttgagttc 120
aaccacacgg gagagctgct ggccacaggt gacaagggcg gccgggtcgt catcttccag 180
cgggaaccag agagtaaaaa tgcgccccac agccagggcg aatacgacgt gtacagcact 240
ttccagagcc acgagccgga gtttgactat ctcaagagcc tggagataga ggagaagatc 300
aacaagatca agtgggtccc acagcagaac gccgccact cactcctgtc caccaacgat 360
aaaactatca aattatggaa gattaccgaa cgagataaaa ggcccgaagg atacaacctg 420
aaggatgaag aggggaaact taaggacctg tccacggtga cgtcactgca ggtgccagtg 480
ctgaagccca tggatctgat ggtggaggtg agccctcgga ggatctttgc caatggccac 540
acctaccaca tcaactccat ctccgtcaac agtgactgcg agacctacat gtcggcggat 600

```

gacctgcgca	tcaacctctg	gcacctggcc	atcacccgaca	ggagcttcaa	catcgtggac	660
atcaagccgg	ccaacatgga	ggaccttacg	gagggtgatca	cagcatctga	gttccatccg	720
caccactgca	acctcttcgt	ctacagcagc	agcaagggct	ccctgcgggt	ctgcgacatg	780
cgggcagctg	ccctgtgtga	caagcattcc	aagctctttg	aagagcctga	ggaccccgagt	840
aaccgctcat	tcttctcgga	aatcatctcc	tccgtgtccg	acgtgaagtt	cagccacagc	900
ggccgctaca	tgctcacccg	ggactacctt	acagtcaagg	tctgggacct	gaacatggag	960
gcaagaccca	tagagacctt	ccaggtccat	gactaccttc	ggagcaagct	ctgttccctg	1020
tacgagaacg	actgcatttt	cgacaagttt	gaatgtgcct	ggaacgggag	cgacagcgtc	1080
atcatgaccg	gggcctacaa	caacttcttc	cgcattgttcg	atcggaacac	caagcgggac	1140
gtgaccctgg	aggcctcgag	ggaaagcagc	aagccccggg	ctgtgctcaa	gccacggcgc	1200
gtgtgcgtgg	ggggcaagcg	ccggcgtgat	gacatcagtg	tggacagctt	ggacttcacc	1260
aagaagatcc	tgcacacggc	ctggcacccg	gctgagaaca	tcattgccat	cgccgccacc	1320
aacaacctgt	acatcttcca	ggacaaggta	aactctgaca	tgcactaggt	atgtgcagtt	1380
cccggccccct	gccaccacagc	ctcatgcaag	tcacccccga	catgaccttc	acgaccgcaa	1440
tgcaaggaggt	ggaagaaagt	cacagcactg	atgaggacag	ctgcagaggt	ggcagtgtgt	1500
ggacacagga	agtttgggcc	ccctccctgc	cccagctttc	ctaggccaga	attgtgtttg	1560
gcagtaattg	tctgtttaaa	aaaataaaaa	ggagaggaag	cgttcaccgc	cgcaaatcat	1620
aaaatggaca	tgactgtgga	gtcttacagt	tcagggt			1657

&lt;210&gt; 15

&lt;211&gt; 2118

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID NO: 1234795CB1

&lt;400&gt; 15

gggaaaagag	gacttattgt	tgtcatggcc	catgagatga	ttggaactca	aattgttact	60
gagaggttgg	tggtctctgt	ggaaagtggg	acggaaaaag	tgctgcta	tgatagccgg	120
ccatttgggt	aatacaatac	atcccacatt	ttggaagcca	ttaatataca	ctgctccaag	180
cttatgaagc	gaaggttgca	acaggacaaa	gtgttaatta	cagagctcat	ccagattcca	240
gcgaaacata	aggttgacat	tgattgcagt	gcagaaggtg	tagtttacga	tcaaagctcc	300
caagatgttg	cctctctctc	ttcagactgt	tttctcactg	tacttctggg	taaactggag	360
aagagcttca	actctgttca	cctgcttgca	ggtgggtttg	ctgagtcttc	tcgttggttc	420
cctggcctct	gtgaaggaaa	atccactcta	gtccctacct	gcatttctca	gccttgctta	480
cctgttgcca	acattggggc	aacccgaatt	cttcccaatc	tttatcttgg	ctgccagcga	540
gatgtcctca	acaaggagct	gatgcagcag	aatgggattg	gttatgtgtt	aaatgccagc	600
aatacctgtc	caaagcctga	ctttatcccc	gagtctcatt	tcctgcgtgt	gcctgtgaat	660
gacagctttt	gtgagaaaat	tttgccgtgg	ttggacaaat	cagtagattt	cattgagaaa	720
gcaaaagcct	ccaatggatg	tgttctagtg	cactgtttag	ctgggatctc	ccgctccgcc	780
accatcgcta	tcgcctacat	catgaagagg	atgcagatgt	ctttagatga	agcttacaga	840
tttgtgaaag	aaaaaagacc	tactatatct	ccaaacttca	attttctggg	ccaactcctg	900
gactatgaga	agaagattaa	gaaccagact	ggagcatcag	ggccaaagag	caaactcaag	960
ctgctgcacc	tggagaagcc	aaatgaacct	gtccctgctg	tctcagaggg	tggacagaaa	1020
agcgagacgc	ccctcagtc	accctgtgcc	gactctgcta	cctcagaggg	agcaggacaa	1080
aggcccgtgc	atcccgccag	cgtgccccagc	gtgccccagc	tgcagccgtc	gctgttagag	1140
gacagcccgc	tggtacaggc	gctcagtggt	ctgcacctgt	ccgcagacag	gctggaagac	1200
agcaataagc	tcaagcgttc	cttctctctg	gatatacaat	cagtttcata	ttcagccagc	1260
atggcagcat	ccttacatgg	cttctctctc	tcagaagatg	ctttggaata	ctacaaacct	1320
tccactactc	tggatgggac	caacaagcta	tgccagttct	cccctgttca	ggaactatcg	1380
gagcagactc	ccgaaaccag	ttccctgata	aggaggaagc	cagcatcccc	aagaagctgc	1440
agaccgccag	gccttcagac	agccagagca	agcgattgca	ttcggtcaga	accagcagca	1500
gtggcaccgc	ccagaggtcc	cttttatctc	cactgcacgc	aagtgggagc	gtggaggaca	1560
attaccacac	cagcttcctt	ttcggccttt	ccaccagcca	gcagcacctc	acgaagtctg	1620
ctggcctggg	cccttaaggg	ctggcactcg	gatatacttg	ccccccagac	ctctaccctt	1680
tcctgacca	gcagctggta	ttttgccaca	gagtcctcac	acttctactc	tgcttcagcc	1740
atctacggag	gcagtgccag	ttactctgcc	tacagctgca	gccagctgcc	cacttgccga	1800
gaccaagtct	attctgtgcg	caggcggcag	aagccaagtg	acagagctga	ctcgcggcgg	1860
agctggcatg	aagagagccc	ctttgaaaag	cagtttaaac	gcagaagctg	ccaaatggaa	1920
tttgagagaga	gcattcatgtc	agagaacagg	tcacgggaag	agctggggaa	agtgggcagt	1980

```

cagtctagct tttcgggagc catggaaatc attgaggtct cctgagaaga aagacacttg 2040
tgacttctat agacaatttt ttttttcttg ttcacaaaaa aattccctgt aaatctgaaa 2100
tatatatatg tacatacc                                     2118

```

&lt;210&gt; 16

&lt;211&gt; 2116

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID NO: 1271505CB1

&lt;400&gt; 16

```

ggaggcggct ccgcgcgcgt ccgctgctag gacccgggca gggctggagc tgggctggga 60
tcccagagctc ggcagcagcg cacggggcgg cccacctgct ggtgccctgg aggctctgag 120
ccccggcggc gcccgggccc acgcggaacg acggggcgag atgcgagcca cccctctggc 180
tgctcctgcg ggttcctgt ccaggaagaa gcggttggag ttggatgaca acttagatac 240
cgagcgtccc gtccagaaac gagctcgaag tgggccccag cccagactgc cccctgcct 300
gttggccctg agcccaccta ctgctccaga tcgtgcaact gctgtggcca ctgcctccc 360
tcttgggccc tatgtcctcc tggagcccga ggaggcggg cgggcctacc aggccctgca 420
ctgccctaca ggcactgagt atacctgcaa ggtgtacccc gtccaggaag ccctggccgt 480
gctggagccc tacgcgcggc tgccccgcga caagcatgtg gctcggccca ctgaggtcct 540
ggctggtacc cagtcctct acgcctttt cactcggacc catggggaca tgcacagcct 600
ggtgcgaacg cgccaccgta tccctgagcc tgaggctgcc gtgctcttcc gccagatggc 660
caccgccctg gcgcactgtc accagcacgg tctggtcctg cgtgatctca agctgtgtcg 720
ctttgtcttc gctgaccgtg agaggaagaa gctggtgctg gagaacctgg aggactcctg 780
cgtgctgact gggccagatg attcctgtg ggacaagcac gcgtgccag cctacgtggg 840
acctgagata ctacgtcac gggcctcata ctcgggcaag gcagccgatg tctggagcct 900
gggcgtggcg ctcttcacca tgctggccgg ccactacccc tccaggact cggagcctgt 960
cctgctcttc ggcaagatcc gccgcggggc ctacgccttg cctgcaggcc tctcgggccc 1020
tgcccgtgt ctggttcgct gcctccttcg tcgggagcca gctgaacggc tcacagccac 1080
aggcatcctc ctgcaccct ggctgcgaca ggaccgatg cccttagccc caaccgatc 1140
ccatctctgg gagctgccc aggtggtccc tgatggactg gggctggacg aagccaggga 1200
agaggaggga gacagagaag tggttctgta tggctaggac caccctacta cacgctcagc 1260
tgccaacagt ggattgagtt tgggggtagc tccaagcctt ctctgcctc tgaactgagc 1320
caaaccttca gtgccttcca gaaggagaa aggcagaagc ctgtgtggag tgtgtgtgt 1380
acacatctgc tttgttccac acacatgcag ttctgtctg ggtgcttatc aggtgccaa 1440
ccctgttctc ggtgctggga gtacagcagt gagcaaagga gacaatatc cctgctcaca 1500
gagatgacaa actggcatcc ttgagctgac aacacttttc catgaccata ggtcactgtc 1560
tacactgggt acactttgta ccagtgtcgg cctccactga tgctggtgct caggcacctc 1620
tgtccaagga caatccctt cacaacaaa ccagctgcct ttgtatcttg taccttttca 1680
gagaaaggga ggtatccctg tgccaaaggc tccaggcctc tccctgcaa ctcaggacct 1740
aagcccagct cactctggga actgtgttcc cagcatctct gtctcttga ttaagagatt 1800
ctccttccag gcctaagcct gggatttggg ccagagataa gaatccaaac tatgaggcta 1860
gttcttgtct aactcaagac tgttctggaa tgagggtcca ggctgtcaa ccatggggct 1920
tctgacctga gcaccaaggt tgagggacag gattaggcag ggtctgtcct gtggccacct 1980
ggaaagtccc aggtgggact cttctgggga cacttggggg ccacaatccc aggtccatac 2040
tctaggtttt ggataccatg agtatgtatg tttacctgtg cctaataaag gagaattatg 2100
aaataaaaaa aaaaaa                                     2116

```

&lt;210&gt; 17

&lt;211&gt; 2897

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID NO: 1385073CB1

&lt;400&gt; 17

```

ctagccgaag cggctgcate tggcgcccg tctgccccgc gtgctcggag cggattctgc 60
ccgccgtccc cggagccctc ggcgccccgc tgagccccgc atcacttcct ccctgtgacc 120
aaccggcgct gcagggttaga gcctggcaat gccgtttggg tgtgtgactc tggcgacaa 180
gaagaactat aaccagccat cggaggtgac tgacagatat gattttgggac aggtcatcaa 240
gactgaggag ttttgtgaaa tcttcggggc caaggacaag acgacaggca agctgcacac 300
ctgcaagaag ttccagaagc gggacggccg caaggtgcgg aaagctgcc aagaacgagat 360
aggcatcctc aagatggtga agcatcccaa catcctacag ctggtggatg tgtttgtgac 420
ccgcaaggag tactttatct tcttgagct ggccacgggg agggaggtgt ttgactggat 480
cctggaccag ggctactact cggagcgaga cacaagcaac gtggtacggc aagtcctgga 540
ggccgtggcc tatttgcact cactcaagat cgtgcacagg aatctcaagc tggagaacct 600
ggtttactac aaccggctga agaactcgaa gattgtcatc agtgacttcc atctggctaa 660
gctagaaaat ggcctcatca aggagccctg tgggaccccc gagtatctgg cccagaggt 720
ggtaggccgg cagcggtatg gacgccctgt ggactgctgg gccattggag tcatcatgta 780
catcctgctt tcaggcaacc cacctttcta tgaggaggtg gaagaagatg attatgagaa 840
ccatgataag aatctcttcc gcaagatcct ggctggtgac tatgagtttg actctccata 900
ttgggatgat atttcgcagg cagccaaaga ctcccatgag aggtgatgg aggtggagca 960
agaccagcgg atcactgcag aagaggccat ctcccatgag tggatttctg gcaatgctgc 1020
ttctgataag aacatcaagg atggtgtctg tgcccagatt gaaaagaact ttgccaggc 1080
caagtggaa aaggtgtcc gagtgaccac cctcatgaaa cggctccggg caccagagca 1140
gtccagcacg gctgcagccc agtcggcctc agccacagac actgccacc cggggctgc 1200
agaccgtagt gccaccccag ccacagatgg aagtgccacc ccagccactg atggcagtg 1260
caccccagcc accgatggaa gcactactcc agccactgat gggagtgtca cccagccac 1320
tgacaggagc gctactccag ccactgatgg gagagccaca ccagccacag aagagagcac 1380
tgtgcccacc acccaaagca gtgccatgct ggccaccaag gcagctgcc cccctgagcc 1440
ggctatggcc cagccggaca gcacagcccc agagggcgcc acagggcagg ctccaccctc 1500
tagtaaaggg gaagaggctg ctggttatgc ccaggagtct caaagggagg agggcagctg 1560
agtaggcagc ctggtgaggg ggggcagggg atgggcagga ggggtggaga gtggatgag 1620
ggcttctcac tgtacataga gtcactggca tgatgccctc gctcccccat gccccacat 1680
cccagtgggg cataactagg ggtcacggga gagcagtctc gtctctctgt tgtatgtgtg 1740
tgagtgggtg gcaggccagt ggcaggggcg gccccagccc ctgcatggat tccttgtggc 1800
ttttctgtct tttgctagct tcaccagttt ctgttccttg tgggatgctg ctctagggat 1860
actcaggggg ctctgtctct ccttccccct cccttcttgc ctaccattc ccctaggcag 1920
gccctgcagg tcccacactc tcccaggccc taaacttggg cggccttgc ctgagagctg 1980
gtcctccagc gaggccctgt cagcggctct aggtcctgc acatgaaggt gtgtgacctg 2040
ggtgtgtggg ctgctctagg agcagataca ggctggtata gaggatgcag aaaggtaggg 2100
cagtatgttt aagtccagac ttggcacatg gctagggata ctgctacta gctgtggagg 2160
tcctcaggag tggagagaat gagtaggagg gcagaagctt ccatttttgt ccttcctaag 2220
accctgttat ttgtgttatt tcctgccttt ccgagtcctg cagtgggctg ccctgtacct 2280
tgaacctcat gagcctctaa gggaaaggag gaacaattag gacgtggcaa tgagacctgg 2340
cagggcagag tacaagccca gcacccagtg tcccagcctt actgggtcct taccctgggc 2400
caaacaggga gggctgatac ctcttgcctc ttccatagat ccacctcct acaatctcag 2460
cccacaagtc ctctccaccc taggggggct gctgcagggc aataactcat aatctgattt 2520
ggaggtttgc cctttacagg ggcagatttt ctgctcagtt caacaatgaa atgaagagga 2580
actccctctt tctacagctc acttctatca gaagcccagg tgcctcagag ccacattgag 2640
ttgctttttc tgggatgagg aagtaggggt aaactcccca gtttcctgag ggaggtcct 2700
gacaggtgcc ctttgtcaga ccctaccaca gcctggatag gcagccacat tggctctgc 2760
ccttgcctcg cactccgtgg tggctctgcc cttctccctg catgcctgtg ggtctgctct 2820
ggtgtgtgaa ggtcgggtgg ttaactgtgt gcctactgaa cctggcaaat aaacatcacc 2880
ctgcaaaagc aaaaaaa
2897

```

<210> 18  
 <211> 839  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID NO: 1606974CB1

<400> 18  
 gtggtgagtt acttggctcg gagcgggcga ggggacgcgt gggcgagcg gggctggcca 60  
 gcctcggccc ccatgacccg ctgtcctgtg ccctttccca gcgatgggcg tgcagcccc 120

```

caacttctcc tgggtgcttc cgggccgggt ggcgggactg gcgctgccgc ggctccccgc 180
ccactaccag ttctgttggt acctgggctg gcggcacctg gtgtccctga cggagcgcg 240
gccccctcac agcgacagct gccccggcct caccctgcac cgctgcgca tccccgactt 300
ctgcccgcgc gcccccgacc agatcgaccg cttcgtgcag atcgtggacg aggccaacgc 360
acggggagag gctgtgggag tgcactgtgc tctgggcttt ggccgcaactg gcaccatgct 420
ggcctgttac ctggtgaagg agcggggctt ggctgcagga gatgccattg ctgaaatccg 480
acgactacga cccggctcca tcgagaccta tgagcaggag aaagcagtct tccagttcta 540
ccagcgaacg aaataagggg ccttagtacc cttctaccag gccctcactc cccttcccca 600
tggtgtcgat ggggccagag atgaaggga gtggactaaa gtattaaacc ctctagctcc 660
cattggctga agacactgaa gtagccacc cctgcaggca ggtcctgatt gaaggggagg 720
cttgactgc tttgttgaat aaatgagttt tacgaaccaa aaaaaaaaaa aaaaaaggcg 780
gccgcaagct tattcccttt agtgagggtt aatttttagct tggcactggc cgctgtata 839

```

&lt;210&gt; 19

&lt;211&gt; 1081

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID NO: 1829744CB1

&lt;400&gt; 19

```

gggaacctca ggcttcagag agccgaaaag ttgggaggcg taaccactta caggccggaa 60
gtgtccgggg tggacgcatt cgggtagccg aagaagtccc aggattgccg aagaagtccc 120
aggatttccg aagcgagccg aagcatcgcg acagttttca gagacagctg atcggttgga 180
gctgttgccg cgagcagtc tggcggcgcc cagagctact acgcccggcg atggcgagga 240
gcccgcggcg gaggtgagg ctctggccgc agcccgggag cggagcagcc gcttcttgag 300
cggcctggag ctggtgaagc aggtgcccga ggcgcgcgtg ttccgtggcc gcttccaggg 360
ccgcgcggcg gtgatcaagc accgcttccc caagggtac cggcaccgcg cgctggaggc 420
gcggcttggc agacggcgga cgggtgcagga ggcccggcg ctcctccgct gtcgcgcgc 480
tggaatatct gcccagttg tcttttttgt ggactatgct tccaactgct tatatatgga 540
agaaattgaa ggctcagtg ctgttcgaga ttatattcag tccactatgg agactgaaaa 600
aactccccag ggtctctcca acttagccaa gacaattggg caggttttgg ctcgaatgca 660
cgatgaagac ctcatctatg gtgatctcac cacctccaac atgctcctga aacccccct 720
ggaacagctg aacattgtgc tcatagactt tgggctgagt ttcatctcag cacttccaga 780
ggataaggga gtagacctct atgtcctgga gaaggccttc ctcatgccc atcccaacac 840
tgaaactgtg tttgaagcct ttctgaagag ctactccacc tcctccaaaa aggccaggcc 900
agtgtcaaaa aaattagatg aagtgcgcct gagaggaaga aagagggtcca tgggtgggta 960
gaagaatgtg tatgacaacc acacacagtg aagctctttt ttcaaagtaa atttgaagaa 1020
atgctacaag tatgagatga gatctaagta aagggtgtta gatattttta aaaaaaaaaa 1080
a 1081

```

&lt;210&gt; 20

&lt;211&gt; 2924

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID NO: 4030831CB1

&lt;400&gt; 20

```

ggcgggggacc cggatgtgtg tgggtggcgcc ggcgaagag cttgtgtgcg gagctgagag 60
gcctatggat gaggaggacg cggcggcccc ggtttgttct catgaacaag atggatgacc 120
tcaacctgca ctaccggttt ctgaattggc gccggcgat cggggagatt cgagagggtc 180
gagctttccg atatcaggag aggttcaaac atatccttgt agatggagat actttaagtt 240
atcatggaaa ctctggtgaa gttggctgct acgtggcttc tcgaccctg accaaggaca 300
gcaattattt tgaggtgtct attgtggaca gtggagtcgg gggcaccatt gctgtggggc 360
tggtccctca gtactacagc ttggatcacc agcctggctg gttgcctgac tctgtagcct 420
accatgctga tgatggcaag ctgtacaatg gccgagccaa gggccgccag tttgggtcaa 480

```

```

agtgcaactc cggggaccgg attggctgtg gcattgagcc tgtgtccttt gatgtgcaga 540
ccgcccagat cttcttcacc aaaaatggga agcgggtggg ctctaccatc atgcccattgt 600
ccccagatgg actgttccca gcagtgggca tgcactccct ggggtgaggag gtgcggctgc 660
acctcaacgc tgagctgggc cgtgaggacg acagcgtcat gatggtggac agttacgagg 720
atgaatgggg ccggctacat gatgtcagag tctgtgggac tctgctggag tacttaggga 780
agggcaaaaa catcgtggat gtggggctgg cccaggcccg gcaccactc agcaccgcga 840
gccactactt cgaggtggag atcgtggacc ctggagagaa atgctacatc gccctggggc 900
tggcacggaa ggactatccc aagaacaggc accctggctg gagcagaggg tctgtggctt 960
atcatgcaga cgatgggaag atcttccatg gcagtgggtg gggggacccc tttgggccac 1020
gctgttataa aggggacatc atgggctgtg gaatcatgtt cccccgggac tacatttttg 1080
acagtgaggg ggacagtgat gacagttgtg acacagtgat cctgtctcog actgcccggg 1140
ccgtccggaa cgtgcggaat gtcatgtacc tgcaccagga aggggaagag gaagaggagg 1200
aagaggaaga ggaagaggat ggggaagaga tagagccgga gcatgagggg aggaagggtg 1260
tggttttctt cactcggaaat ggcaagatca ttgggaagaa ggatgctgtt gttccttctg 1320
gaggcttctt ccccaccatt ctcccctcca gacctgctcc ttctccctgc tcaccctctg 1440
accccttgag tggctagggc ctgacttccc agaggcttcg tttaccagc agggccctgg 1500
ctgggccagg caccagttc cccactggct caggccctg tcacgcttct ctgtgccac 1560
aggtgtgtag tcaactctgcc actgttgta gtccctgggc ctgagtcctt ggttgagacag 1620
gtttctgacc tgggtgctgcc gtgtgtatgt ggggtgtccc actcgctttg gtcagtgggc 1680
gaatggaccc aaagaatggt gttggtatgt ggtgtgggtg agaggcgtga gcaccctatc 1740
ttctgggtcc ccctttccct caccggccct gtgtgggtg gagggtagag tagacagccc cctcccctac 1800
tcagctgcta ttcgggcatg atgctttgta gagggtagag tagacagccc cctcccctac 1800
tcaccatggt atttctcctt gaattcctct ttctgtttt ctttctctgt tgtgtgaacc 1860
agttgctgct gtcatacccc tggcaggggc aggggacctc tctttggta tctctgtcct 1920
ttcactggct gctgccccag gaagactcct ctaggctctc catctttccc ttgagagctg 1980
gtccccacc ccaacctgct caggcaccac agaggatcta ggtctctggc tccccatacc 2040
tggaccaca tgggtgggtg cctgttgcat gtttaagaga gaggggctgt gaggtgacag 2100
ggcactaggg ccttcaactc tttctcccct tccatcctt ctttaccagt gccaccatg 2160
tccctagctc ccgggtattg gggctgaggc tctggggcct gtctccctgc cagcgtgagg 2220
gcaagacccc agagccttag ctgagcaagc ccagaggggc agcgtggccc ctccctcccc 2280
ttttcctgcc ccgtcccatg cctcagcttg ctgcttgtgc cagttgcctg tttcgcttca 2340
gtgtttgatt ctagcactta catgtgtcct cccaccaag ccctctatct ccttctaate 2400
cttcaacccc tggccccctc ccgtaacag tgacttttcc agggaggaag aggcagcagg 2460
agctgttggc cttggtttgc acagagcggg tagggctgta gggaaagcgg gtgagctgtt 2520
gtgctgctgg gcctcccttt ggccctcgct tcccacccta cgatgtatga aatgtatgta 2580
cagaccagag atgtttatac agccgataaa gatggagttt ccgtatttat cagtatggcc 2640
ggaaccagga gcctttctag tccactgggc taggaacagg actgctggat gggggcagcc 2700
gaaggcagct tgctcatggg gagatgtgga ccaatgttgg gccagggatg ggaatcatat 2760
gttccatggg cctggctaca ggccctgagca cagatacgtc ccctgggaga tgaggctttg 2820
accttctgt gaataagtgt tgactccaat ttccgctaaa gtttatagaa attctttatt 2880
attagacaaa aatagactct cttttttccc ctaaaaaaaa aaaa 2924

```

&lt;210&gt; 21

&lt;211&gt; 2781

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID NO: 5039718CB1

&lt;400&gt; 21

```

cggctcgagg cgggtccagg actgtccgcc ggggtgaggg aaggggccgt gcccggtgcc 60
agcccagggt ctcgcgccct ggctccatgg ccctgggtcac agtgagccgt tcgcccccg 120
gcagcggcgc ctccacgccc gtggggccct gggaccaggc ggtccagcga aggagtcgac 180
tccagcgaag gcagagcttt gcggtgctcc gtggggctgt cctgggactg caggatggag 240
gggacaatga tgatgcagca gaggccagtt ctgagccaac agagaaggcc ccgagtgagg 300
aggagctcca cggggaccag acagacttcg ggcaaggatc ccagagtccc cagaagcagg 360
aggagcagag gcagcacctg cacctcatgg tacagctgct gaggcgcag gatgacatcc 420
ggctggcagc ccagctggag gcaccccggt ctccccgggt ccgtacctg ctggtagttt 480
ctacacgaga aggagaaggt ctgagccagg atgagcgggt cctcctgggc gtggatttcc 540
ctgacagcag ctccccccag tgcaccctgg gcctgggtctt gcccctctgg agtgacaccc 600

```

```

aggtgtactt agatggagac gggggcttca gcgtgacgtc tgggtgggcaa agccggatct 660
tcaagcccat ctccatccag accatgtggg ccacactcca ggtattgcac caagcatgtg 720
aggcagctct aggcagcggc cttgtaccgg gtggcagtg cctcacctgg gccagccact 780
accaggagag actgaactcc gaacagagct gcctcaatga gtggacggct atggccgacc 840
tggagtctct gcggcctccc agtgccgagc ctggcgggtc ctcagaacag gagcagatgg 900
agcaggcgat ccgtgctgag ctgtggaaag tgttggatgt cagtgaacctg gagagtgtca 960
cttccaaaaga gatccgccag gctctggagc tgcgcctggg gctccccctc cagcagtacc 1020
gtgacttcat cgacaaccag atgctgctgc tgggtggcaca gcgggaccga gcctcccgca 1080
tcttccccca cctctacctg ggctcagagt ggaacgcagc aaacctggag gagctgcaga 1140
ggaacagggg caccacatc ttgaacatgg cccgggagat tgacaacttc taccctgagc 1200
gcttcaccta ccacaatgtg cgcctctggg atgaggagtc ggcccagctg ctgccgact 1260
ggaaggagac gcaccgcttc attgaggctg caagagcaca gggcaccac gtgctggtcc 1320
actgcaagat gggcgctcagc cgctcagcgg ccacagtgtt ggcctatgcc atgaagcagt 1380
acgaatgcag cctggagcag gccctgcgcc acgtgcagga gctccggccc atcgcccgcc 1440
ccaacctggg ctctctgccc cagctgcaga tctaccaggg catcctgacg gccagccgcc 1500
agagccatgt ctgggagcag aaagtgggtg gggctctccc agaggagcac ccagcccctg 1560
aagtctctac accattccca cctcttccgc cagaacctga ggggtggggg gaggagaagg 1620
ttgtaggcat ggaagagagc caggcagccc cgaaagaaga gcctgggcca cggccacgta 1680
taaacctccg aggggtcatg aggtccatca gtcttctgga gccctccttg gagctggaga 1740
gcacctcaga gaccagtgc atgccagagg tcttctcttc ccacgagtct tcacatgaag 1800
agcctctgca gcccttccca cagcttgcaa ggaccaaggg aggccagcag gtggacaggg 1860
ggcctcagcc tgccctgaag tcccgcagc cagtggttac cctccagggc agtgccgtgg 1920
tggccaaccg gaccagggcc ttccaggagc aggagcaggg gcaggggagc gggcagggag 1980
agccctgcat ttcctctacg cccagggttc ggaaggtggg gagacaggcc agcgtgcatg 2040
acagtggaga ggagggcgag gcctgagccc tcacacatgc ccacgctccc ctgacactga 2100
agaggatcca caactccttg gagaaacacc ctcacgtctg ttgccgcaca cattcctctc 2160
agctccgccc cataccgctc actacagcct cactcccac ccctgtcact acggcctcac 2220
ctcccacccc tgctactaca gcctcacctc ctacagcctt aagtcccagg cccatgtctg 2280
cctgtccaag ggctcaagac tttctaactg ggatgtggta gagggactga aggtaccttt 2340
gggggcaaca gcaccctagt ttcatctca actctagccc tgcacactca cctgtggcac 2400
ggaatgaaaa cagagcttcc cgtgcaaaaa gggtcacgcc tcccaccccc gccccctccc 2460
tgcacctcct gtctctctcc agttcattcc tggaaaccagc caggccaggc aaccagtggc 2520
cccaaaggc aggcaggatc ctcaggcccc agccgcggga ggctggaagg gctggcagat 2580
cgcttccctc atccacctcc accggtccag gtctttgctg ctgtccccag acctcctgtg 2640
acaccacgcc agatcacagg gcaccaggcc agagatagtc ttctttttgt cctttctggc 2700
ctctggctag tcagtttttc atagccttac agtatctggc tttgtactga gaaataaaac 2760
acattttcat aaaaaaaaaa a 2781

```

&lt;210&gt; 22

&lt;211&gt; 754

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID NO: 5595281CB1

&lt;400&gt; 22

```

gttctctgtc tctcacttc cctctgtggc tctgctcaga actggcggtt tttcccagct 60
ccttgcccag accaatactt ccatgctgtc ttcaagccct gcttctctga catctcccag 120
cccagatggg gagaacccat gtaagaaggt ccactgggct tctgggagga gaaggacatc 180
atccacagac tcagagtcca agtcccaccc ggactcctcc aagataccca ggtcccggag 240
accagccgc ctgacagtga agtatgaccg gggccagctc cagcgctggc tggagatgga 300
gcaatgggtg gatgctcaag ttcaggagct cttccaggat caagcaaccc cttctgagcc 360
tgagattgac ctggaagctc tcatggatct atccacagag gagcagaaga ctcagctgga 420
ggccattctt gggaaactgcc cccgccccac agaggctttt atctctgagc tgctcagtea 480
actcaagaaa ctccggagac tcagccggcc tcagaaataa gcctgagaga ccatcttttag 540
cagcctcagc actgccaggc ctgccctgaa actccagatc ctggctaaga gggaaatagc 600
tccttgggac acaaaacaaga aatgtggaca aggagggaca tttgcatact cctactgtct 660
gtgtggctac agctagtttc tgtcagctgg gctctctggg agaaagctgg ctgttgtcca 720
atgccttctt tggcagccaa gtggataaaa cctt 754

```